(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 May 2004 (13.05.2004)

PCT

(10) International Publication Number
WO 2004/039825 A2

(51) International Patent Classification[7]: C07H 21/00

(21) International Application Number:
PCT/DK2003/000739

(22) International Filing Date: 30 October 2003 (30.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PA 2002 01652	30 October 2002 (30.10.2002)	DK
60/422,167	30 October 2002 (30.10.2002)	US
PA 2002 01955	19 December 2002 (19.12.2002)	DK
60/434,425	19 December 2002 (19.12.2002)	US
PA 2003 01064	11 July 2003 (11.07.2003)	DK
60/486,199	11 July 2003 (11.07.2003)	US

(71) Applicant *(for all designated States except US)*: **NUEVOLUTION A/S** [DK/DK]; Rønnegade 8, 5th floor, DK-2100 Copenhagen Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FRESKGÅRD, Per-Ola** [SE/SE]; Örtugsgatan 50, S-603 79 Norrköping (SE). **FRANCH, Thomas** [DK/DK]; Humlebækgade 14, st, tv., DK-2200 København N (DK). **GOULIAEV, Alex, Haahr** [DK/DK]; Brøndsted 223, DK-3670 Veksø Sjælland (DK). **LUNDORF, Mikkel, Dybro** [DK/DK]; Viborggade 13, st. tv., DK-2100 Cpèemjagen Ø (DK). **FELDING, Jakob** [DK/DK]; Ordruphøjvej 24, 1., DK-2920 Charlottenlund (DK). **OLSEN, Eva, Kampmann** [DK/DK]; Vingetoften 17, DK-2730 Helev (DK). **HOLTMANN, Anette** [DK/DK]; Langekærvej 42, DK-2750 Ballerup (DK). **JAKOBSEN, Søren, Nyboe** [DK/DK]; Frederiksvej 36, 4. tv., DK-2000_ Frederiksberg (DK). **SAMS, Christian** [DK/DK]; Jakob Dannefaerds Vej 4a 1, DK-1973 Frederiksberg C (DK). **GLAD, Sanne, Schrøder** [DK/DK]; Viggo Barfoeds Allé 59, DK-2750 Ballerup (DK). **JENSEN, Klm, Birkebæk** [DK/DK]; Voldumvej 30C, DK-2610 Rødovre (DK). **PEDERSEN, Henrik** [DK/DK]; Frodesvej 24, DK-2880 Bagsvaerd (DK).

(74) Agent: **HOIBERG A/S**; Store Kongensgade 59 A, DK-1264 Copenhagen K. (DK).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ENZYMATIC ENCODING

(57) Abstract: Disclosed is a method for obtaining a bifunctional complex comprising a display molecule part and a coding part, wherein a nascent bifunctional complex comprising a chemical reaction site and a priming site for enzymatic addition of a tag is reacted at the chemical reaction site with one or more reactants, and provided with respective tag(s) identifying the reactants(s) at the priming site using one or more enzymes.

WO 2004/039825 A2

Enzymatic encoding

This application claims the benefit of U.S. provisional application Serial No. 60/422,167, filed October 30, 2002; U.S. provisional application Serial No. 60/434,425, filed December 19, 2002, and U.S. provisional application Serial No. 60/486,199, filed July 11, 2003, which are hereby incorporated by reference in their entirety. All patent and non-patent references cited in these patent applications, or in the present application, are hereby incorporated by reference in their entirety.

Technical Field of the Invention

The present invention relates to a method for obtaining a bifunctional complex comprising display molecule part and a coding part. The invention also relates to a method for generation of a library of bifunctional complexes, a method for identifying a display molecule having a preselected property.

Background

Approaches have been developed that allow the synthetic encoding of polypeptides and other biochemical polymers. An example of this approach is disclosed in US 5,723,598, which pertains to the generation of a library of bifunctional molecules. One part of the bifunctional complex is the polypeptide and the other part is an identifier oligonucleotide comprising a sequence of nucleotides which encodes and identifies the amino acids that have participated in the formation of the polypeptide. Following the generation of the library of the bifunctional molecules, a partitioning with respect to affinity towards a target is conducted and the identifier oligonucleotide part of the bifunctional molecule is amplified by means of PCR. Eventually, the PCR amplicons are sequenced and decoded for identification of the polypeptides that have affinity towards the target. The library of bifunctional complexes is produced by a method commonly known as split-and-mix. The method implies that a linker molecule is divided into spatial separate compartments and reacted with a specific amino acid precursor at one terminus in each compartment and appended a nucleic acid tag which codes for this specific amino acid precursor at the other terminus by an orthogonal chemical reaction. Subsequently, the content of the various compartments are collected (mixed) and then again split into a number of compartments for a new round of alternating reaction with amino acid precursor and nucleotide tag. The split-and-mix method is continued until the desired length of polypeptide is reached.

This prior art method is constrained in its application because there must be compatible chemistries between the two alternating synthesis procedures for adding a chemical unit as compared to that for adding a nucleotide or oligonucleotide sequence. According to the prior art, the problem of synthesis compatibility is solved by the correct choice of compatible protecting groups as the alternating polymers are synthesised, and by the correct choice of methods for deprotection of one growing polymer selectively while the other growing polymer remains blocked.

Halpin and Harbury have in WO 00/23458 suggested another approach, wherein the molecules formed are not only identified but also directed by the nucleic acid tag. The approach is also based on the split-and-mix strategy to obtain combinatorial libraries using two or more synthetic steps. A plurality of nucleic acid templates are used, each having at one end a chemical reactive site and dispersed throughout the stand a plurality of codon regions, each of said codon regions in turn specifying different codons. The templates are separated by hybridisation of the codons to an immobilised probe and subsequently each of the strands is reacted at the chemical reaction sites with specific selected reagents. Subsequently, all the strands are pooled and subjected to a second partitioning based on a second codon region. The split-and-mix method is conducted an appropriate number of times to produce a library of typically between 10^3 and 10^8 different compounds. The method has the disadvantage that a large number of nucleic acid templates must be provided. In the event a final library of 10^6 different compounds is desired, a total of 10^8 nucleic acid templates must be synthesised. The synthesis is generally cumbersome and expensive because the nucleic acids templates must be of a considerable length to secure a sufficient hybridisation between the codon region and the probe.

In WO 02/074929 a method is disclosed for the synthesis of chemical compounds. The compounds are synthesised by initial contacting a transfer unit comprising an anti-codon and a reactive unit with a template having a reactive unit associated therewith under conditions allowing for hybridisation of the anti-codon to the template and subsequently reacting the reactive units. Also this method suffers from the disadvantage that a large number of nucleic acid templates initially must be provided.

The prior art methods using templates suffer from the disadvantage that encoding is dependent upon the recognition between the anti-codon and the template. The hybridisation between two oligonucleotides can occur in the event there is a sufficient complementarity between these. Occasionally, the hybridisation will occur even though a complete match between the oligonucleotides is not present. The effect is, in the event a plurality of transfer units are present then sometimes the codon sequence of the template does not correspond to the reactive unit actually reacted. This undesired effect is even more pronounced when the formation of library is intended because a plurality of templates and building blocks are supposed to find each other in the reaction media. When the hybridisation step is not completely correct, molecules will be generated that are encoded by the incorrect codons on the template. This will have two major effects on the selection process performed on the library. First, templates with a codon combination encoding for binding ligands will be lost in the selection process. Secondly, and may be more important, templates with a codon combination encoding for non-binding ligands will be enriched.

In an aspect of the present invention it is an object to provide a non-template dependent method for obtaining an encoded molecule, said method allowing for versatile chemistries to be applied in the formation of the encoded molecule, because the application of compatible orthogonal protection groups in the alternating formation of the encoded molecule and oligonucleotide tag can be avoided. The present invention in a preferred aspect intends to improve on the error prone hybridisation method previous suggested in the codon recognition process. Furthermore, it is an object of the invention to reduce non-specific reaction products formed. Thus, in an aspect of the present invention, the present method has an inherent proof-reading facility securing that the phenotype is accurately encoded by the genotype.

Summary of the Invention

The present invention relates to a method for obtaining a bifunctional complex comprising a display molecule part and a coding part, wherein a nascent bifunctional complex comprising a chemical reaction site and a priming site for enzymatic addition of a tag is reacted at the chemical reaction site with one or more reactants, and provided with respective tag(s) identifying the reactant(s) at the priming site using one or more enzymes.

Enzymes are in general substrate specific, entailing that the enzymatic addition of a tag to the priming site is not likely to interfere with the display molecule being formed. Thus, the application of protection groups on the coding part as well as the nascent display molecule can be avoided for this reason. However, it may be desired for other reasons to protect the growing display molecule. Enzymes are available having an activity in aqueous and organic media. The vast majority of enzymes, however, have a higher activity in an aqueous media compared to an organic media. Therefore, prior to or subsequent to the providing of the tag it may be desired to change the media in order to obtain applicable conditions for the reaction of the reactant at the chemical reaction site.

Generally, the display molecule part is formed by more than a single round of reaction between one or more reactants and the chemical reaction site. In a certain aspect of the invention, the nascent bifunctional complex reacted with one or more reactants and provided with respective tag(s) is reacted further one or more times with one or more reactant(s) and is provided with respective identifying tag(s) to produce a reaction product as one part of the bifunctional complex and an identifying part comprising tags which codes for the identity of the reactants which have participated in the formation of the reaction product.

In a certain aspect of the invention, a round or cycle of reaction implies that a single reactant is reacted with the chemical reaction site and that a respective tag identifying the reactant is provided at the priming site for enzymatic addition. In another aspect of the invention, a round of reaction implies that multiple reactants are reacted at the chemical reaction site and that tags identifying one or more, but not necessarily all, reactants are provided at the priming site for enzymatic addition. The reaction at the chemical reaction site and the addition of tags may occur in any order, i.e. the reaction may occur subsequent to, simultaneously with, or previous to the tag addition. The choice of order may among other things be dependent on the enzyme type, the reaction conditions, and the type of reactant.

The nascent bifunctional complex comprises a chemical reaction site and a priming site for enzymatic addition of a tag. Optionally, the nascent bifunctional complex also comprises a linking moiety, which connects the chemical reaction site with the priming site.

The linking moiety may serve various purposes, such as distancing the priming site from the chemical reaction site sufficient from each other to allow an enzyme to perform the tag addition and provide for a hybridisation region. In an aspect of the invention, the linking moiety is a nucleic acid sequence. The length of the oligonucleotide is preferably suitable for hybridisation with a complementing oligonucleotide, i.e. the number of nucleotides in the linking moiety is suitably 8 or above. In a certain embodiment, the linking moiety is attached to the chemical reaction site via a spacer comprising a selectively cleavable linker to enable a detachment of the display molecule from the coding part in a step subsequent to the formation of the final bifunctional complex. A nascent bifunctional complex is also referred to as a growing complex and specifies an initial or intermediate complex to be processed according to the method of the present invention. An intermediate complex designates an initial complex that has been subjected to one or more rounds of reactant reaction and tag addition.

The chemical reaction site may comprise a single or multiple reactive groups capable of reacting with one or more reactants. In a certain aspect the chemical reaction site comprises a scaffold having one or more reactive groups attached. Examples of suitable reactive groups include amine, carboxylic acid, thio, aldehyde, and hydroxyl groups. Examples of scaffolds include benzodiazepines, steroids, hydantiones, piperasines, diketopiperasines, morpholines, tropanes, cumarines, qinolines, indoles, furans, pyrroles, oxazoles, amino acid precursors, and thiazoles. Furthermore, the reactive groups of the chemical reaction site may be in a pro-form that has to be activated before a reaction with the reactant can take place. As an example, the reactive groups can be protected with a suitable group, which needs to be removed before a reaction with the reactant can proceed. A display molecule in the present description with claims indicates a chemical reaction site that has been reacted with one or more reactants.

The reactants of the present invention include free reactants as well as reactants which comprises a functional entity and a nucleic acid sequence. The free reactant participates in the reaction with the chemical reaction site and may give rise to a chemical structure of the final display molecule. A functional entity attached to a nucleic acid may be referred to herein as a building block and specifies a chemical entity in which the functional entity is capable of being reacted at the chemical reaction site. In a certain aspect of the invention, the functional entity is detached from the nucleic acid part and

transferred to the chemical reaction site. The oligonucleotide of the building block may or may not hold information as to the identity of the functional entity. In a certain embodiment of the present invention, the reactant is a building block comprising an oligonucleotide sufficient complementary to the linking moiety to allow for hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity. The free reactant is generally not attached to a nucleic acid unless a nucleic acid component is intended in the final display molecule. The free reactant may have any chemical structure and preferably comprises a reactive group or a precursor therefore, which will enable a reaction with a chemical reaction site. Examples of reactive groups include hydroxyl groups, carboxylic acid groups, thiols, isocyanates, amines, esters, and thioesters. Optionally, a further reactant occurs to mediate a connection between the free reactant and the chemical reaction site. The functional entity of a building block resembles the free reactant as far as the requirement for reaction with the chemical reaction site concerns. In addition, however, it is in most instances necessary to cleave the connection between the functional entity and the nucleic acid following the reaction. Optionally, the reaction and cleavage may occur in a single step. Various types of building blocks are disclosed in detail below. In a certain aspect of the invention, the free reactant or the functional entity do not include a nucleotide.

The coding part of the nascent bifunctional complex is formed by addition of at least one tag to a priming site using one or more enzymes. Further tags may be attached to a previous tag so as to produce a linear or branched identifier. As long as at least one tag of the identifier is attached by an enzymatic catalysed reaction, further tags may be provided using chemical means or enzymatic means at the discretion of the experimenter. In a certain embodiment of the invention, all tags are provided using an enzymatic catalysed reaction. A tag suitably comprises recognition units, i.e. units which may be recognized by recognition groups. The recognition unit possess an ability to carry information so as to identify a reactant. A variety of different kinds of recognition exist in nature. Examples are antibodies, which recognise an epitope, proteins which recognise another protein, mRNA which recognise a protein, and oligonucleotides which recognise complementing oligonucleotide sequences. Generally, it is preferred that the tag is a sequence of nucleotides.

The coding part of the bifunctional complex is in a preferred aspect of the invention amplifiable. The capability of being amplified allows for the use of a low amount of bi-

functional complex during a selection process. In the event, the tag is a protein, the protein may be amplified by attaching the mRNA which has encoded the synthesis thereof, generating the cDNA from the mRNA and subjecting said mRNA to a translation system. Such system is disclosed in WO 98/31700, the content of which is incorporated herein by reference. An alternative method for amplifying a protein tag is to use phage displayed proteins. In general, however, the tag is a sequence of nucleotides, which may be amplified using standard techniques like PCR. When two or more tags are present in a linear identifying oligonucleotide, said oligonucleotide generally consist of a certain kind of backbone structure, so as to allow an enzyme to recognise the oligonucleotide as substrate. As an example the back bone structure may be DNA or RNA.

The priming site of a nascent bifunctional complex is capable of receiving a tag. The chemical identity of the priming site depends among other things on the type of tag and the particular enzyme used. In the event the tag is a polynucleotide, the priming site generally comprises a 3'-OH or 5'-phosphate group of a receiving nucleotide, or functional derivatives of such groups. Enzymes which may be used for enzymatic addition of a tag to the priming site include an enzyme selected from polymerase, ligase, and recombinase, and a combination of these enzymes.

The reaction between the chemical reaction site and the one or more reactants may take place under suitable conditions that favours the reaction. In some aspects of the invention, the reaction is conducted under hybridisation conditions, i.e. an annealing between two complementing oligonucleotides remains during the reaction conditions. In other aspects of the invention, the reaction is conducted under denaturing conditions to allow for suitable condition for the reaction to occur. In the event, the coding part of the growing complex comprises an oligonucleotide; said oligonucleotide is in an aspect of the invention in a double stranded form during the reaction to reduce the likelihood of side reactions between components of the oligonucleotide and reactants.

The tag identifying a reactant can be added to the priming site using any appropriate enzyme. In a certain embodiment, a tag is provided at the priming site of the nascent bifunctional complex utilizing an enzymatic extension reaction. The extension reaction may be performed by a polymerase or a ligase or a combination thereof. The extension using a polymerase is suitably conducted using an anti-tag oligonucleotide as template.

The anti-tag oligonucleotide is annealed at the 3' end of the oligonucleotide part of the nascent bifunctional complex with a single stranded overhang comprising an anti-codon, which identifies the reactant. The anti-codon of the anti-tag can be transcribed to the identifier part using a polymerase and a mixture of dNTPs. Alternatively, a ligase is used for the addition of the tag using one or more oligonucleotides as substrates. The ligation can be performed in a single stranded or a double stranded state depending on the enzyme used. In general it is preferred to ligate in a double stranded state, i.e. oligonucleotides to be ligated together are kept together by a complementing oligonucleotide, which complements the ends of the two oligonucleotides.

Examples of suitable enzymes include DNA polymerase, RNA polymerase, Reverse Transcriptase, DNA ligase, RNA ligase, Taq DNA polymerase, Pfu polymerase, Vent polymerase, HIV-1 Reverse Transcriptase, Klenow fragment, or any other enzyme that will catalyze the incorporation of complementing elements such as mono-, di- or polynucleotides. Other types of polymerases that allow mismatch extension could also be used, such for example DNA polymerase η (Washington et al., (2001) JBC 276: 2263-2266), DNA polymerase ι (Vaisman et al., (2001) JBC 276: 30615-30622), or any other enzyme that allow extension of mismatched annealed base pairs. In another aspect, when ligases are used, suitable examples include Taq DNA ligase, T4 DNA ligase, T4 RNA ligase, T7 DNA ligase, and *E. coli* DNA ligase. The choice of the ligase depends to a certain degree on the design of the ends to be joined together. Thus, if the ends are blunt, T4 RNA ligase may be preferred, while a Taq DNA ligase may be preferred for a sticky end ligation, i.e. a ligation in which an overhang on each end is a complement to each other.

The tag added to the priming site of the nascent bifunctional complex holds information as to the reactant. In the present invention with claims, the information relating to the reactant will be termed codon. Apart from a combination of the nucleotides coding for the identity of the reactant, a tag may comprise further nucleotides. In a certain aspect of the invention, a tag comprises a framing sequence. The framing sequence may serve various purposes, such as an annealing region for anti-tags and/or as a sequence informative of the point in time of the synthesis history the associated reactant has reacted.

The association between the codon and the identity of the reactant may vary dependent on the desired output. In a certain embodiment, the codon is used to code for several different reactants. In a subsequent identification step, the structure of the display molecule can be deduced taking advantage of the knowledge of the different attachment chemistries, steric hindrance, deprotection of orthogonal protection groups, etc. In another embodiment, the same codon is used for a group of reactants having a common property, such as a lipophilic nature, molecular weight, a certain attachment chemistry, etc. In a preferred embodiment however, the codon is unique, i.e. a similar combination of nucleotides does not identify another reactant. In a practical approach, for a specific reactant, only a single combination of nucleotides is used. In some aspects of the invention, it may be advantageous to use several different codons for the same reactant. The two or more codons identifying the same reactant may carry further information related to different reaction conditions. In another aspect of the invention, a single codon specifies two or more reactants.

In one aspect of the invention, each bifunctional complex is prepared by simultaneous or sequentially tagging and reaction of reactant as illustrated in the scheme below:

Capital letters represent reactant or chemical reaction site. Lower case letters represent tags.

A scaffold "X" is linked to a tag "x". A reactant is linked to "X" e.g. "A" and so is a tag for that fragment e.g. "a". Suitably, the tag is unique.

The coding part of the eventually formed bifunctional complex will contain all the codons. The sequence of each of the codons is used to decipher the structure of the reactants that have participated in the formation of the displayed molecule, i.e. the reaction product. The order of the codons can also be used to determine the order of incorporation of the reactants. This may be of particular interest when a linear polymer is formed, because the exact sequence of the polymer can be determined by decoding the encoding sequence. Usually, to facilitate the decoding step, a constant or binding region is transferred to the bifunctional complex together with the codon. The constant

region may contain information about the position of the related reactant in the synthesis pathway of the display molecule.

The invention also relates to a method for identifying a display molecule having a preselected property, comprising the steps of: subjecting the library produced according to the method indicated above to a condition, wherein a display molecule or a subset of display molecules having a predetermined property is partitioned from the remainder of the library, and identifying the display molecule(s) having a preselected function by decoding the coding part of the complex.

The above method, generally referred to as selection, involves that a library is subjected to a condition in order to select display molecules having a property which is responsive to this condition. The condition may involve the exposure of the library to a target. The bifunctional complexes having an affinity towards this target may be partitioned form the remainder of the library by removing non-binding complexes and subsequent eluting under more stringent conditions the complexes that have bound to the target. Alternatively, the coding part of the bifunctional complex can be cleaved from the display molecule after the removal of non-binding complexes and the coding part may be recovered and decoded to identify the display molecule.

It is possible to perform a single or several rounds of selection against a specific target with a subsequently amplification of the selected variants. These obtained variants are then separately tested in a suitable assay. The selection condition can be stringent and specific to obtain binding molecules in one selection rounds. It may be advantageously to perform the method using a single round of selection because the number and diversity of the potential binders are larger compared to procedures using further selections where potential binders may be lost. In another embodiment the selection procedure involves several round of selection using increasing stringency conditions. Between each selection an amplification of the selected complex may be desirable.

The coding part can be amplified using PCR with primers generating two unique cut-sites. These cut-sites can be used for multimerization of the coding region by cloning into a suitable vector for sequencing. This approach will allow simultaneously sequencing of many encoding regions. Alternatively, the PCR product is directly cloned into a suitable vector using for example TA cloning. In still another approach the

identity of the display molecule is established by applying the PCR product to a suitable microarray.

It is within the capability of the skilled person in the art to construct the desired design of an oligonucleotide. When a specific annealing temperature is desired it is a standard procedure to suggest appropriate compositions of nucleic acid monomers and the length thereof. The construction of an appropriate design may be assisted by software, such as Vector NTI Suite or the public database at the Internet address http://www.nwfsc.noaa.gov/protocols/oligoTMcalc.html. The conditions which allow hybridisation of two oligonucleotides are influenced by a number of factors including temperature, salt concentration, type of buffer, and acidity. It is within the capabilities of the person skilled in the art to select appropriate conditions to ensure that the contacting between two oligonucleotides is performed at hybridisation conditions. The temperature at which two single stranded oligonucleotides forms a duplex is referred to as the annealing temperature or the melting temperature. The melting curve is usually not sharp indicating that the annealing occurs over a temperature range.

The present invention may be conducted in two basic modes. A first mode uses a reactant in which a codon or anti-codon covalently is connected to the functional entity which it identifies. A second mode uses a reactant which is not covalently attached to a codon or anti-codon. The tag is provided at the priming site of the bifunctional complex by an entity separate from the reactant. When more than a single round is carried out, the first and the second mode can be combined in any order. When a library of different bifunctional complexes is to be generated, the two modes are conducted in accordance with two different approaches. A library produced using the first mode can be conducted in a single vessel, which herein will be referred to as a one-pot synthesis, whereas a library produced according to the second mode requires a split-and-mix synthesis, i.e. the reaction and tag addition must be carried out in separate compartments for each complex. In a certain embodiment of the invention, one or more tags coding for two or more reactants, respectively, are provided prior to or subsequent to the reaction involving the two or more reactants and the chemical reaction site.

Mode 1:

The present invention relates in a first mode to a method for encoding the identity of a chemical entity transferred to a bifunctional complex, said method comprising the steps of

a) providing a nascent bifunctional complex comprising a reactive group and an oligonucleotide identifier region,

b) providing a building block comprising an oligonucleotide sufficient complementary to the identifier region to allow for hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity,

c) mixing the nascent bifunctional complex and the building block under hybridisation conditions to form a hybridisation product,

d) transferring the functional entity of the building block to the nascent bifunctional complex through a reaction involving the reactive group of the nascent bifunctional complex, and

e) enzymatically extending the oligonucleotide identifier region to obtain a codon attached to the bifunctional complex having received the chemical entity.

The method of the invention involves the incorporation of a codon for the functional entity transferred to the complex. The incorporation of the codon is performed by extending over an anticodon of the building block using an appropriate enzyme, i.e. an enzyme active on nucleic acids. The transcription of the encoding region can be accomplished by an enzyme, such as a polymerase or a ligase. In general, it is preferred to use enzymes which are specific toward the substrate and the end-product to obtain an as accurate as possible transcription of the anti-codon. A high degree of specificity is generally available for nucleic acid active enzymes because a non-specific activity could destroy the ability of the living cells to survive. Especially preferred enzymes according to the present invention are polymerases with proof-reading activity for accurate encoding but preservation of the upstream nucleobases.

The enzymatic extension may occur subsequent to or simultaneously with the transfer of the functional entity or even prior to the transfer. However, in general it is preferred to perform the extension step subsequent to the transfer step to avoid any possible interaction between the enzyme and the functional entity.

As the enzyme will perform extension only when the identifier region and the complementing identifier region has hybridised to each other to form a double helix, it is secured that the functional entity and the reactive group has been in close proximity when the complex is provided with a codon. Compared to the hybridisation method previously suggested, the present invention has the advantage that complexes provided with functional entities through a non-directed reaction will not be provided with a codon. Thus, false positive molecules may easily be detected due to the absence of a codon.

The invention also relates to a method for obtaining a bifunctional complex composed of a display molecule part and a coding part, wherein the method for encoding the identity of a chemical entity transferred to a bifunctional complex further comprises step f) separating the components of the hybridisation product and recovering the complex.

The invention may be performed by transferring only a single functional entity and the corresponding codon to the nascent bifunctional complex. However, in general it is preferred to build a display molecule composed of two of more functional entities. Thus, in a preferred aspect of the invention a method is devised for obtaining a bifunctional complex composed of a display molecule part and a coding part, said display molecule part being the reaction product of functional entities and the reactive group of the initial complex, wherein steps c) to f) are repeated as appropriate. In the final cycle of the preparation of the bifunctional complex, step f) may be dispensed with, notably in cases in which a double stranded identifier oligonucleotide is obtained because a double stranded nucleic acid usually is more stable compared to a corresponding single stranded oligonucleotide. The identifier oligonucleotide may also become double stranded by an extension process in which a primer is annealed to the 3'end of the oligonucleotide and extended using a suitable polymerase. The double strandness may be an advantage during subsequent selection processes because a single stranded nucleic acid may perform interactions with a biological target, in a way similar to aptamers. In the repetition of the cycle, the produced bifunctional complex in a previous cycle, i.e. a nascent bifunctional complex that has received a functional entity and a codon, is used as the nascent bifunctional complex in the next cycle of functional entity transfer and codon incorporation.

The oligonucleotides used according to the present method are of a reasonable extent. Thus, the long pre-made templates suggested in the prior art (in WO 00/23458 it is suggested to use oligonucleotides of at least 220 and preferably 420 nucleotides) are generally avoided.

The invention also relates to a method for generating a library of bifunctional complexes, comprising the steps of:

a) providing one or more different nascent bifunctional complexes comprising a reactive group and an oligonucleotide identifier region,

b) providing a plurality of different building blocks, each comprising an oligonucleotide sufficient complementary to an identifier region to allow for hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity,

c) mixing nascent bifunctional complexes and plurality of building blocks under hybridisation conditions to form hybridisation products,

d) transferring functional entities of the building blocks to the nascent bifunctional complexes through a reaction involving the reactive group of the nascent bifunctional complex,

e) enzymatically extending the oligonucleotide identifier regions to obtain codons attached to the bifunctional complexes having received the chemical entities,

f) separating the components of the hybridisation products and recovering the complexes,

g) repeating steps c) to f) one or more times, as appropriate.

A disadvantage associated with the hybridisation technique suggested in the prior art becomes apparent when the formation of libraries are considered. Even though two double stranded oligonucleotides have the same number of nucleotides it is by no means ensured that they will possess the same melting temperature. This is at least partly due to the fact that different number of hydrogen bondings are involved for different base pairs (the C-G pair involves three hydrogen bondings and the A-T base pair involves two hydrogen bondings). Thus, establishing a temperature for the annealing of various building blocks to a template will be a compromise between avoiding mismatching and ensuring sufficient annealing. The present invention aims at avoiding this disadvantage by providing, in a preferred embodiment of the invention, an identifier region having a similar affinity towards all building blocks.

In the event, more than one identifier sequence is used, e.g. when more than one kind of reactive group or scaffolds are present, a building block occasionally may be mis-annealed thereto. However, the transferred functional entity will actually be correctly encoded on the complex through the extension process. This approach resembles the arrangement Nature is using: Allowing mis-incorporation of bases at the DNA level (compare to mismatch annealing of building blocks) to obtain diversification but insisting on correct encoding for the phenotype (compare to the extension of the right codon on the complex).

The annealing between the identifier and the building block can either be a random process or be guided by the sequences in the identifier region and the complementing identifier region. A random process can be achieved by using the same sequence in all identifier regions and the complementing identifier regions. Thus a mixture of identifiers and building blocks will anneal randomly or simi-randomly and create unique combinations of functional entities. Alternatively, a random or simi-random process can be achieved by using universal bases at positions of the building block opposing nucleobases of the identifier that codes for the identity of a particular scaffold or reactive group. The sequences of the identifier oligonucleotides and the building block oligonucleotides may be optimized such that it is assured that the sequences in a library involved in the annealing process will assemble at an equal degree of annealing regardless of which functional entity that is attached to the building block. Thus, there will be no or diminished bias in the selection procedure due to different annealing properties for specific building blocks. In addition, the similarities in the annealing process in each annealing step and for each hybridisation product in a library will make sure the functional entity is presented equally for the reactive group/scaffold. This will provide optimal conditions for the transfer step.

The nascent bifunctional complex comprises an oligonucleotide identifier region and a reactive group. The reactive group may be connected to the oligonucleotide through a cleavable linker allowing for the separation of the final reaction product from the oligonucleotide. A single reactive group may be present or multiple reactive groups may be present as a part of a scaffold. The scaffold may be attached to the oligonucleotide through a cleavable linker to allow for subsequent separation of the reacted scaffold. The reactive groups may be selected from any groups capable of receiving a functional entity. Examples of suitable reactive groups include amine,

carboxylic, thio, and hydroxyl groups. Furthermore, the reactive group of the nascent bifunctional complex may be in a pro-form that has to be activated before the method of the invention is initiated. A nascent bifunctional complex is also referred to as a growing complex and specifies an initial or intermediate complex to be further processed according to the present invention.

The number of nucleotides in the identifier region of the identifier molecule is determined from how strong and specific the annealing should be between the identifier and building block. A stronger and more specific annealing process is generally obtained with a longer nucleotide sequence. Normally about 10 – 20 nucleotides is sufficient to achieve specific and efficient annealing. However, in some aspects of the invention the range can be from 2 – 1000, most preferably between 15 – 30 nucleotides.

The identifier region may in certain embodiments comprise information about the identity of the reactive group or the scaffold of the nascent bifunctional complex. Such scaffold codon is generally at a position distanced from the scaffold to allow for the formation of a stable double helix at the part comprising the functional entity to be transferred and the scaffold. The scaffold codon may have any length but is generally selected with the same length as the codons specifying the functional entities. The rear part of the identifier region is generally provided with a constant or binding sequence. The binding sequence when annealed to a suitable part of the building block provides for a substrate for the enzyme to perform the extension.

The building block comprises an oligonucleotide sufficient complementary to at least a part of the identifier region to allow for hybridisation. The oligonucleotide of the building block may not completely be complementary to the identifier, that is, one or more mismatches may be allowed but it must be assured that the building block is able to anneal to the identifier region. For the sake of simplicity, the part of the building block oligonucleotide capable of annealing to the identifier will be referred to as the complementing identifier region. In the present description with claims, the term hybridisation is to be understood as the process of attaching two single stranded oligonucleotides to each other such that a hybridisation product is formed.

The building block comprises also an anticodon region made of oligonucleotides. The anti-codon identifies the identity of the functional entity of the building block. In a certain embodiment, the same anticodon is used to code for several different functional entities. In a subsequent identification step, the structure of the display molecule can be deduced taking advantage of the knowledge of different attachment chemistries, steric hindrance, deprotection of orthogonal protection groups, etc. In another embodiment, the same anti-codon is used for a group of function entities having a common property, such as a lipophilic nature, a certain attachment chemistry etc. In a preferred embodiment, however, the anti-codon is unique i.e. a similar combination of nucleotides does not appear on another building block carrying another functional entity. In a practical approach, for a specific functional entity, only a single combination of nucleotides is used. In some aspects of the invention, it may be advantageous to use several anti-codons for the same functional entity, much in the same way as Nature uses up to six different anti-codons for a single amino acid. The two or more anti-codons identifying the same functional entity may carry further information related to different reaction conditions.

The individual anti-codons may be distinguished from another anti-codon in the library by only a single nucleotide. However, to facilitate a subsequent decoding process it is in general desired to have two or more mismatches between a particular anticodon and any other anti-codon appearing on the various building blocks. As an example, if a codon/anticodon length of 5 nucleotides is selected, more than 100 nucleotide combinations exist in which two or more mismatches appear. For a certain number of nucleotides in the codon, it is generally desired to optimize the number of mismatches between a particular codon/anticodon relative to any other codon/anticodon appearing in the library.

The coupling of the functional entity to the complementary identifier region can be done with suitable coupling reactions. Any coupling reaction or combination of such reactions known in the art can be used as appropriate as readily recognized by those skilled in the art. The functional entity linked to the complementary identifier region is a molecule, which preferably comprises at least one reactive group that allows linkage to the reactive group of the identifier.

The sequence of the anticodon identifies the functional entity attached in the same building block. This anticodon sequence is either directly included in the building block sequence or is attached to a pre-existing building block using a polymerase or a ligase for example. In a certain embodiment, as disclosed in detail in example 7, complementing identifier regions, termed carrier oligos in the example, are initially loaded with the various functional entities. Each of the loaded carrier oligoes is subsequently ligated to an anti-codon oligo using a splint oligo to assemble the two oligonucleotides. The ligation reaction serves to connect the functional entity to be transferred with an anticodon specifying the structure of the functional entity. The anti-codon oligo may be designed in various ways. Normally, a region that allows for the annealing of the splint is included in the design. However, some ligases like the T4 RNA ligase, does not require a stretch of double stranded DNA. Therefore, the splint and the part of the anti-codon oligo annealing to the splint can be dispensed with in some embodiments. In the event the identifier region comprises a codon coding for the identity of the scaffold, the anti-codon oligo comprises a stretch of universal bases, like inosines. The universal bases may be dispensed with if a region complementing a binding region on the identifier region is included downstream. The latter embodiment normally will entail that a part of the identifier loops out. The complementing binding region is normally selected such that a polymerase is capable of recognizing a formed double helix with a binding region of the nascent bifunctional molecule as a substrate. The anti-codon is suitably positioned at the 5' side of the complementing binding region so it can be transferred to the nascent complex by an extension reaction. Suitably, the complementing binding region is designed such that it is possible to identify the position of the particular codon in the sequence of codons appearing on the eventual bifunctional complex.

The anticodon sequence is transcribed to the identifier through an extension process to form the codon on the identifier molecule. This may be carried out by any state of the art method including, but not limited to, a polymerase extension reaction. A polymerase extension reaction usually requires the presence of sufficient polymerase activity together with each of the four natural nucleotide tri-phosphates (ATP, CTP, GTP, and TTP) in a suitable buffer. Thus, the sequence of a particular anticodon is only transferred to the identifier as a codon when the building block and the identifier molecule has annealed and allow reaction to take place between the functional entity and the recipient reactive group.

The four natural nucleotides can encode for 4^N variants where N is the length of the codon. For example, if the unique codon is 5 nucleotides in length, the number of possible encoding for different functional entities is 1024. The codons can also be design using a sub-set of the four natural nucleotides in each position. This can be useful in combination with the use of universal nucleobases. The anticodon in each building block is coding for the functional entity in the same building block. This sequence may in an aspect of the invention be incorporated by PCR of the complementing identifier region with a functional entity primer and an anticodon primer.

The functional entity of the building block serves the function of being a precursor for the structural entity eventually incorporated into the displayed molecule. Therefore, when in the present application with claims it is stated that a functional entity is transferred to a nascent bifunctional complex it is to be understood that not necessarily all the atoms of the original functional entity is to be found in the eventually formed display molecule. Also, as a consequence of the reactions involved in the connection, the structure of the functional entity can be changed when it appears on the nascent display molecule. Especially, the cleavage resulting in the release of the functional entity may generate a reactive group which in a subsequent step can participate in the formation of a connection between a nascent display molecule and a functional entity.

The functional entity of the building block preferably comprises at least one reactive group capable of participating in a reaction which results in a connection between the functional entity of the building block and the identifier carrying the reactive group. The number of reactive groups which appear on the functional entity is suitably one to ten. A functional entity featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas functional entities having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. Two or more reactive groups intended for the formation of connections, are typically present on scaffolds. A scaffold is a core structure, which forms the basis for the creation of multiple variants. The variant forms of the scaffold are typically formed through reaction of reactive groups of the scaffold with reactive groups of other functional entities, optionally mediated by fill-in groups or catalysts. The functional entities to be connected to the scaffold may contain one, two or several reactive groups able to form connections. Examples of scaffold include steroids, hydantions, benzodiazepines, etc.

The reactive group of the building block may be capable of forming a direct connection to a reactive group of the identifier or the reactive group of the building block may be capable of forming a connection to a reactive group of the identifier through a bridging fill-in group. It is to be understood that not all the atoms of a reactive group are necessarily maintained in the connection formed. Rather, the reactive groups are to be regarded as precursors for the structure of the connection.

After or simultaneously with the formation of the connection a cleavage is performed to transfer the functional entity to the identifier. The cleavage can be performed in any appropriate way. In an aspect of the invention the cleavage involves usage of a reagent or and enzyme. The cleavage results in a transfer of the functional entity to the nascent bifunctional complex or in a transfer of the complex to the functional entity of the building block. In some cases it may be advantageous to introduce new chemical groups as a consequence of linker cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it is desirable that no trace of the linker remains after the cleavage.

In another aspect, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the functional entity of the building block or the nascent display molecule is a leaving group of the reaction. In some aspects of the invention, it is preferred to design the system such that the connection and the cleavage occur simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above. In other aspects of the invention, it is preferred to conduct separate cross-linkage and cleavage steps because the stepwise approach allows for mastering each sub steps and for a reduction in the likelihood for non-specific transfer.

Preferably, at least one linker remains intact after the cleavage step. The at least one linker will link the nascent display molecule to the encoding region. In case the method essentially involves the transfer of functional entities to a scaffold or an evolving polymer, the eventually scaffolded molecule or the polymer may be attached with a selectively cleavable linker. The selectively cleavable linker is designed such that it is not

cleaved under conditions which result in a transfer of the functional entity to the nascent template-directed molecule.

The cleavable linkers may be selected from a large plethora of chemical structures. Examples of linkers includes, but are not limited to, linkers having an enzymatic cleavage site, linkers comprising a chemical degradable component, and linkers cleavable by electromagnetic radiation. Cleavable linkers of particular interest are currently linkers that can be cleaved by light. A suitable example includes an *o*-nitro benzyl group positioned between the display molecule and the identifier region.

The building blocks used in the method according to the present invention may be designed in accordance with the particular entities involved in the building block. As an example, the anti-codon may be attached to the complementing identifier region with a polyethylene glycol (PEG) linker and the functional entity may be directly attached to said complementing identifier region. In another and preferred example, the anti-codon, complementing identifier region and the functional entity is a contiguous linear oligonucleotide. In a certain embodiment of the invention, the building block is designed such that a part of the identifier loops out. The loop out of the identifier usually occurs because the building block oligo does not anneal to the entire length of the identifier. Usually, the building block is designed such that it is able to anneal to at least the identifier region of the bifunctional complex and to a binding region at the rear part of the identifier. The complementing identifier region and the anticodon may be directly connected through a single linkage, connected through a PEG linker of a suitable length, or a sequence of nucleobases which may or may not comprise nucleobases complementing the various codons and binding region on the identifier. In a certain embodiment of the invention, the building block is designed only to anneal to a binding region, usually at an end of the identifier opposing the end having attached the display molecule. In an aspect of the invention the building block and/or the nascent identifier are composed of two or more separate nucleotides, which are able to hybridise to each other to form the hybridisation complex. The gaps between the oligonucleotides may be filled with suitable nucleotide using an appropriate enzyme activity, such as a polymerase and a ligase, to produce a coherent identifier and or building block.

The attachment of the functional entity to the complementing identifier region is usually conducted through a linker. Preferably the linker connects the functional entity with the

complementing identifier region at a terminal nucleotide or a nucleotide 1 or two nucleotides down the oligonucleotide. The attachment of the functional entity can be at any entity available for attachment, i.e. the functional entity can be attached to a nucleotide of the oligonucleotide at the nucleobase, or the back bone. In general, it is preferred to attach the functional entity at the phosphor of the internucleoside linkage or at the nucleobase.

In a certain aspect of the invention, the reactive group of the functional entity is attached to the linker oligonucleotide. The reactive group is preferably of a type which is able to create a connection to the nascent display molecule by either direct reaction between the respective reactive groups or by using a suitable fill-in group. The reactive group coupling the functional entity with the linker is preferably cleaved simultaneously with the establishment of the connection. The functional entity may in some cases contain a second reactive group able to be involved in the formation of a connection in a subsequent cycle. The second reactive group may be of a type which needs activation before it is capable of participating in the formation of a connection.

In the event two or more functional entities are to be transferred to the complex, the codons may be separated by a constant region or a binding region. One function of the binding region may be to establish a platform at which the polymerase can bind. Depending on the encoded molecule formed, the identifier may comprise further codons, such as 3, 4, 5, or more codons. Each of the further codons may be separated by a suitable binding region. Preferably, all or at least a majority of the codons of the identifier are separated from a neighbouring codon by a binding sequence. The binding region may have any suitable number of nucleotides, e.g. 1 to 20.

The binding region, if present, may serve various purposes besides serving as a substrate for an enzyme. In one setup of the invention, the binding region identifies the position of the codon. Usually, the binding region either upstream or downstream of a codon comprises information which allows determination of the position of the codon. In another setup, the binding regions have alternating sequences, allowing for addition of building blocks from two pools in the formation of the library. Moreover, the binding region may adjust the annealing temperature to a desired level.

A binding region with high affinity can be provided by incorporation of one or more nucleobases forming three hydrogen bonds to a cognate nucleobase. Examples of nucleobases having this property are guanine and cytosine. Alternatively, or in addition, the binding region may be subjected to backbone modification. Several backbone modifications provides for higher affinity, such as 2'-O-methyl substitution of the ribose moiety, peptide nucleic acids (PNA), and 2'-4' O-methylene cyclisation of the ribose moiety, also referred to as LNA (Locked Nucleic Acid).

The identifier may comprise flanking regions around the codons. The flanking region can encompass a signal group, such as a flourophor or a radio active group to allow for detection of the presence or absence of a complex or the flanking region may comprise a label that may be detected, such as biotin. When the identifier comprises a biotin moiety, the identifier may easily be recovered.

The flanking regions can also serve as priming sites for amplification reactions, such as PCR. Usually, the last cycle in the formation of the bifunctional complex includes the incorporation of a priming site. The identifier region of the bifunctional complex is usually used for another priming site, thereby allowing for PCR amplification of the coding region of the bifunctional complex.

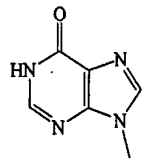
It is to be understood that when the term identifier is used in the present description and claims, the identifier may be in the sense or the anti-sense format, i.e. the identifier can comprise a sequence of codons which actually codes for the molecule or can be a sequence complementary thereto. Moreover, the identifier may be single-stranded or double-stranded, as appropriate.

The design of the part of the complementing identifier region or the building block oligonucleotide in general which comprises one or more anti-codons preceding the active anti-codon can be random or simi-random and one or more mismatches with the identifier region may be allowed. However, especially when a library is contemplated, it may be advantageous to incorporate in a region complementing a preceding codon one or more non-specific base-pairing nucleobases. Non-specific base-pairing nucleobases are bases which, when attached to a backbone, are able to pair with at least two of the five naturally occurring nucleobases (C, T, G, A, and U). Preferably, the base pairing between the two or more natural nucleobases and the non-specifically

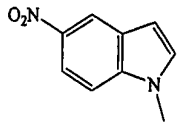
base-pairing nucleobase occur essentially iso-enegically, i.e. the bonds formed have a strength of the same order. The term "non-specifically base-pairing nucleobase" is used herein interchangeably with the term "universal base".

In natural tRNA, the nucleobase inosine is found. Inosine has the ability to hybridise non-specifically with three of the nucleobases, i.e. cytosine, thymine, and adenine. Inosine and examples of other synthetic compounds having the same ability of non-specifically base-pairing with natural nucleobases are depicted below

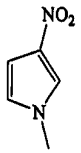
Examples of Universal Bases:



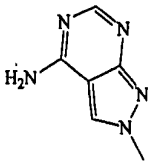
Inosine



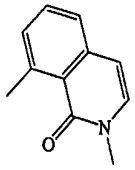
5-Nitroindole



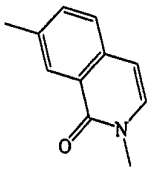
3-Nitropyrrole



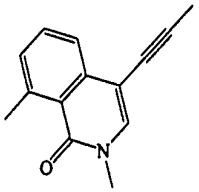
N^8-8aza-7deazaadenine



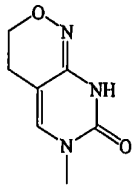
MICS



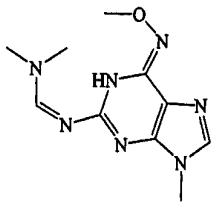
5MICS



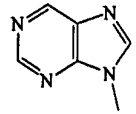
PIM



dP



dK



Nebularine

The use of universal bases in the present method has an advantage in the generation of a library because the nucleobases of previously transferred codons can be matched with universal bases on the complementing region of the building block. The

complementing of a spent codon with a sequence of universal bases allows for the use of the same building block for a variety of growing bifunctional complexes.

The encoding by extension principle can also be used using a three-strand procedure. Each step involves a library of assembly platform molecules hybridised to a functional entity carrier (Figure 7). The assembly platform comprise a fixed sequence (complementing identifier region) that binds equally well to all or a subset of identifier molecule through the identifier region. Alternatively, this complementing identifier sequence can also be random or simi-random to increase the diversity of the library as this would allow for the use of different scaffold molecules. The assembly platform also contains a unique anticodon region with a specific sequence. This specific sequence will anneal to the unique codon region in the carrier, thus forming a building block in which the transferable functional entity is coupled to a unique anti-codon by hybridisation. The sequence of the unique anticodon and the unique anticodon region is linked allowing a direct coupling between these two sequences. This coupling is for example obtained when the assembly platform is synthesized.

The unique anticodon can either be identical to the unique anticodon region or a shorter or longer sequence. However, a prerequisite though is that these two sequences (the unique anticodon and the unique anticodon region) are linked to each other, e.g. through the complementing identifier region and, optionally, the connection region. The sequence of the unique anticodon can be used to decode the unique anticodon region. This will obtain the unique codon region which codes for the functional entity. The connecting region is optionally a sequence that can be varied to obtain optimal reactivity between functional entity and the attachment entity. If polymers are created using this system, the connecting region could be extended through the assembling cycles.

The formation of identifier-displayed molecules by the three-strand assembly principle is performed in sequential steps. Each individual step involves annealing of the carrier and the identifier molecules to the assembly platform. After the annealing step, two important events take place: 1) the reaction between the attachment entity and the functional entity to accomplish transfer of the functional entity to the identifier molecule, and 2) the extension of the unique codon sequence into the identifier molecule using the unique anticodon sequence on the assembly platform as the reading sequence.

The formation of a library of bifunctional complexes according to the invention can be performed using a solid support for the platform molecule as shown in Fig. 9 and 10. This allow a sequential transfer where each library of assembly platform molecules, with different addition of the non-coding region and complementing binding region dependent of which specific step, is immobilized in separate vials and a library of identifier and building block molecules is supplied. After the annealing- reaction/transfer-extension steps, the library is removed (*e.g.* with elevated temperature) and transferred to another vial with an immobilized assembly platform library (with an additional non-coding and complementing binding region) to allow the next step in the process.

Mode 2:

The present invention discloses in a second mode of the invention, a method for generating a complex comprising a display molecule part and a coding part, wherein a nascent bifunctional complex comprising a chemical reaction site and a priming site for enzymatic addition of a tag is reacted at the chemical reaction site with one or more reactants and provided at the priming site with respective tags identifying the one or more reactants using one or more enzymes.

The lack of a covalent link between the reactive part and the coding part of the building block implies that a library is to be produced by a split-and-mix strategy. In a first step a nascent bifunctional complex is dispensed in one or more separate compartment and subsequently exposed to a reactant in each compartment, which reacts at the chemical reaction site, and an agent which provides the tag identifying said reactant at the priming site. The agent providing the tag includes an enzyme and a substrate therefore. In a certain embodiment of the invention, the tag is provided by extending over an anticodon using a polymerase. In another embodiment of the invention, the tag is provided at the priming site by ligation of a codon oligonucleotide, which holds information as to the identity of the reactant.

When the enzyme is a polymerase, the substrate is usually a blend of triphosphate nucleotides selected from the group comprising dATP, dGTP, dTTP, dCTP, rATP, rGTP, rTTP, rCTP, rUTP. Substrates for ligases are oligo- and polynucleotides, i.e. nucleic acids comprising two or more nucleotides. An enzymatic ligation may be performed in a single or double stranded fashion. When a single stranded ligation is per-

formed, a 3' OH group of a first nucleic acid is ligated to a 5' phosphate group of a second nucleic acid. A double stranded ligation uses a third oligonucleotide complementing a part of the 3' end and 5' end of the first and second nucleic acid to assist in the ligation. Generally, it is preferred to perform a double stranded ligation.

In some embodiments of the invention, a combination of polymerase transcription and ligational coupling is used. As an example, a gap in an otherwise double stranded nucleic acid may be filled-in by a polymerase and a ligase can ligate the extension product to the upstream oligonucleotide to produce a wholly double stranded nucleic acid.

Mode 2 is conducted in separate compartments for each reaction, as discussed above. Thus, the addition of a tag occurs without competing nucleic acids present and the likelihood of cross-encoding is reduced considerable. The enzymatic addition of a tag may occur prior to, subsequent to, or simultaneous with the reaction. In some aspects of the invention, it is preferred to add the tag to the nascent bifunctional complex prior to the reaction, because it may be preferable to apply conditions for the reaction which are different form the conditions used by the enzyme. Generally, enzyme reactions are conducted in aqueous media, whereas the reaction between the reactant and the chemical reaction site for certain reactions is favoured by an organic solvent. An appropriate approach to obtain suitable condition for both reactions is to conduct the enzyme reaction in an aqueous media, lyophilize and subsequent dissolve or disperse in a media suitable of the reaction at the chemical reactive site to take place. In an alternative approach, the lyophilization step may be dispensed with as the appropriate reaction condition can be obtained by adding a solvent to the aqueous media. The solvent may be miscible with the aqueous media to produce a homogeneous reaction media or immiscible to produce a bi-phasic media.

The reactant according to the second mode may be a free reactant or a zipper building block. A free reactant is not attached to a code identifying another part of the reactant. In most cases, a free reactant comprises a chemical structure comprising one, two or more reactive groups, which can react with the chemical reaction site. A zipper building block is a functional entity which is attached to a chemical entity that binds in the vicinity of the chemical reaction site. The binding chemical entity may be an oligonucleotide which hybridises to a linking moiety of the nascent bifunctional complex prior to the

reaction. The hybridisation event will increase the proximity between the functional entity and the chemical reaction site, thereby reducing the possibility of side reactions and promote the reaction due to a high local concentration.

The nascent bifunctional complex is constructed having the encoding method in mind. Thus, if a polymerase is used for the encoding, a region of hybridisation is usually provided in the linker moiety. The region of hybridisation will allow for a binding region of a complementing oligonucleotide comprising an anti-codon to hybridise to the nascent bifunctional complex. The binding region serves as a binding site for a polymerase, which then may produce an extension product using the anti-codon oligonucleotide as template. When a ligase is used for the encoding, the priming site of the nascent bifunctional complex comprises one or more nucleotides which the ligase may consider as a substrate. In a single stranded ligation an oligonucleotide present in the media and bearing information as to the identity of the reactive group will be ligated to the nascent bifunctional molecule. A double stranded ligation requires the priming site of the nascent bifunctional complex to be able to hybridise to a complementing oligonucleotide prior to ligation. Suitably, the priming site comprises one, two, or more nucleotides, to which a complementing oligonucleotide can hybridise. The complementing oligonucleotide hybridise in the other end to the codon oligonucleotide, which holds the information of a particular reactant.

The linker moiety of the nascent bifunctional complex may comprise information relating to the identity of the chemical reaction site. In an applicable approach, the linker moiety comprises a codon informative of the identity of the chemical reaction site.

The oligonucleotides bearing the information on the pertinent reactant, may, apart from the combination of nucleotides identifying the reactant, comprise flanking regions. The flanking regions may serve as binding regions capable of hybridising to the nascent bifunctional complex. The binding region may be designed so as to hybridise promiscuous to more than a single nascent bifunctional complex. Alternatively, the binding region on the coding oligonucleotide is capable of being ligated to a binding region the nascent bifunctional complex using a splint oligonucleotide as mediator.

The invention may be performed by reacting a single reactant with the nascent bifunctional complex and add the corresponding tag. However, in general it is preferred to

build a display molecule comprising the reaction product of two of more reactants. Thus, in a certain aspect of the invention a method is devised for obtaining a bifunctional complex composed of a display molecule part and a coding part, said display molecule part being the reaction product of reactants and the chemical reaction site of the initial complex. In an aspect of the invention, two alternating parallel syntheses are performed so that the tag is enzymatical linked to the nascent bifunctional complex in parallel with a reaction between a chemical reaction site and a reactant. In each round the addition of the tag is followed or preceded by a reaction between reactant and the chemical reaction site. In each subsequent round of parallel syntheses the reaction product of the previous reactions serves as the chemical reaction site and the last-incorporated tag provides for a priming site which allows for the enzymatical addition a tag. In other aspects of the invention, two or more tags are provided prior to or subsequent to reaction with the respective reactants.

The coding part comprising all the tags may be transformed to a double stranded form by an extension process in which a primer is annealed to the 3' end of the oligonucleotide and extended using a suitable polymerase. The double strandness may be an advantage during subsequent selection processes because a single stranded nucleic acid may perform interactions with a biological target in a way similar to aptamers.

In a certain aspect of mode 2 a method is devised for generating a library of bifunctional complexes comprising a display molecule part and a coding part. The method comprises the steps of providing in separate compartments nascent bifunctional complexes, each comprising a chemical reaction site and a priming site for enzymatic addition of a tag and performing in any order reaction in each compartment between the chemical reaction site and one or more reactants, and addition of one or more respective tags identifying the one or more reactants at the priming site using one or more enzymes.

The nascent bifunctional complexes in each compartment may be identical or different. In the event the nascent bifunctional complex differs at the chemical reaction site, the nascent bifunctional complex suitable comprises a codon identifying the structure of the chemical reaction site. Similar, the reactants applied in each compartment may be

identical or different as the case may be. Also, the reaction conditions in each compartment may be similar or different.

Usually, it is desired to react the complex with more than a single reactant. In a certain aspect of the invention, the content of two or more compartments are pooled together and subsequently split into an array of compartments for a new round of reaction. Thus, in any round subsequent to the first round, the end product of a preceding round of reaction is used as the nascent bifunctional complex to obtain a library of bifunctional complexes, in which each member of the library comprises a reagent specific reaction product and respective tags which codes for the identity of each of the reactants that have participated in the formation of the reaction product. Between each round of reaction the content of the compartments is in an aspect of the invention mixed together and split into compartments again. In other aspects of the invention the content of a compartment is after having received a codon but before a reaction has occurred divided into further compartments in which a further codon is received and a reaction occurs with the two reactants that have been encoded. In another aspect of the invention, more than two codons are encoded before a reaction between chemical reaction site and reactants are allowed to take place. In the alternative, two or more reactions are allowed to occur before an encoding with the respective tags is initiated.

The individual codons may be distinguished from another codon in the library by only a single nucleotide. However, to facilitate a subsequent decoding process it is in general desired to have two or more differences between a particular codon and any other codon. As an example, if a codon/anticodon length of 5 nucleotides is selected, more than 100 nucleotide combinations exist in which two or more differences appear. For a certain number of nucleotides in the codon, it is generally desired to optimize the number of differences between a particular codon/anticodon relative to any other codon/anticodon appearing in the library. An oligonucleotide codon may comprise any suitable number of nucleotides, such as from 2 to 100, 3 to 50, 4 to 20 or 5 to 15 nucleotides.

The reactant can be a free reactant or a zipper building block. The reactant serves the function of being a precursor for the structural entity eventually incorporated in to the displayed molecule part. There structure of a reactant may after reaction with a chemical reaction site become changed in a subsequent round. In the event the reactant is a

zipper building block, a cleavage of the linkage between the functional entity and the oligonucleotide is normally conducted after reaction. An exception is in the final round, in which the cleavage can be dispensed with. The cleavage can occur subsequent to or simultaneously with the reaction with the chemical reaction site. The cleavage may generate a reactive group which in a subsequent step can participate in the formation of a connection between the nascent display molecule and a reactant.

The free reactant or the functional entity of the zipper building block preferably comprises at least one reactive group capable of participating in a reaction which results in a connection to the chemical reaction site of the nascent bifunctional molecule. The number of reactive groups which appear on the free reactant and the functional entity is suitably one to ten. A free reactant or a functional entity featuring only one reactive group is used i.a. in the end positions of polymers or scaffolds, whereas functional entities having two reactive groups are suitable for the formation of the body part of a polymer or scaffolds capable of being reacted further. Two or more reactive groups intended for the formation of connections, are typically present on scaffolds. A scaffold is a core structure, which forms the basis for the creation of multiple variants. The variant forms of the scaffold are typically formed through reaction of reactive groups of the scaffold with reactive groups of other reactants, optionally mediated by fill-in groups or catalysts. The functional entities or free reactants to be connected to the scaffold may contain one, two or several reactive groups able to form connections. Examples of scaffolds include steroids, hydantions, benzodiazepines, etc.

The reactive group of the free reactant or the functional entity attached to a nucleic acid comprising a zipper region, i.e. a region promiscuously binding to a linking moiety of the nascent bifunctional complex, may be capable of forming a direct connection to a reactive groups of the chemical reactive site or the reactant may be capable of forming a connection to a reactive group of the chemical reactive site through a bridging fill-in group. It is to be understood that not all the atoms of the reactive groups are necessarily maintained in the connection formed. Rather the reactive groups are to be regarded as precursors for the structure of the connection.

When a zipper building block is used, a cleavage may be performed after or simultaneously with the formation of the connection between the chemical reaction site and the functional entity. The cleavage can be performed in any appropriate way. In an aspect

of the invention the cleavage involves usage of a reagent or enzyme. The cleavage results in a transfer of the functional entity to the nascent bifunctional complex or in a transfer of the complex to the functional entity of the zipper building block. In some cases it may be advantageous to introduce new chemical groups as consequence of the cleavage. The new chemical groups may be used for further reaction in a subsequent cycle, either directly or after having been activated. In other cases it s desirable that no trace of the linker remains after the cleavage. In some aspects of the invention it may not be desired to cleave on or more chemical bonds. As an example, it may be desirable to maintain the connection between the zipper domain and the functional entity in the last round.

In some aspects of the invention, the connection and the cleavage is conducted as a simultaneous reaction, i.e. either the functional entity of the zipper building block or the chemical reactive site of the nascent bifunctional complex is a leaving group of the reaction. In some aspects of the invention, it is preferred to design the system such that the cleavage occurs simultaneously because this will reduce the number of steps and the complexity. The simultaneous connection and cleavage can also be designed such that either no trace of the linker remains or such that a new chemical group for further reaction is introduced, as described above. In other aspects of the invention, it is preferred to conduct separate cross-linking and cleavage steps because the stepwise approach allows for mastering each sub step and for a reduction of the likelihood of non-specific transfer.

The attachment of the functional entity to the oligonucleotide comprising a zipping domain is usually conducted through a linker. Preferably the linker connects the functional entity with the oligonucleotide at a terminal nucleotide or a nucleotide 1 or two nucleotides down the oligonucleotide. The attachment of the functional entity can be at any entity available for attachment, i.e. the functional entity can be attached to a nucleotide of the oligonucleotide at the nucleobase, or the back bone. In general, it is preferred to attach the functional entity at the phosphor of the internucleoside linkage or at the nucleobase.

In a certain aspect of the invention, the reactive group of the functional entity is attached to the oligonucleotide, optionally through a suitable spacer. The reactive group is preferably of a type which is able to create a connection to the nascent display mole-

cule by either direct reaction between the respective reactive groups or by using a suitable fill-in group. The reactive group coupling the functional entity with the oligonucleotide is preferably cleaved simultaneously with the establishment of the connection. The functional entity may in some cases contain a second reactive group able to be involved in the formation of a connection in a subsequent cycle. The second reactive group may be of a type which needs activation before it is capable of participating in the formation of a connection.

Preferably at least one linker remains intact after the cleavage step. The at least one linker will link the display molecule to the coding part, i.e. the part comprising the one or more tags identifying the various reactant that have participated in the formation of the display molecule. It may be desired to connect the display molecule part to the coding part of the bifunctional complex through a space comprising a selectively cleavable linker. The selectively cleavable linker is designed such that it is not cleaved under conditions which result in a transfer of a function entity to the chemical reaction site.

The cleavable linkers may be selected from a large plethora of chemical structures. Examples of linkers includes, but are not limited to, linkers having an enzymatic cleavage site, linkers comprising a chemical degradable component, and linkers cleavable by electromagnetic radiation. Cleavable linkers of particular interest are currently linkers that can be cleaved by light. A suitable example includes an o-nitro benzyl group positioned between the display molecule and the coding part of the bifunctional complex.

In the event two or more reactants are reacted with the chemical reactive site, the codons of the coding part may be separated by a constant region or a binding region. One function of the binding region may be to establish a platform at which an enzyme, such as polymerase or ligase can recognise as a substrate. Depending on the encoded molecule formed, the identifier may comprise further codons, such as 3, 4, 5, or more codons. Each of the further codons may be separated by a suitable binding region. Preferably, all or at least a majority of the codons of the identifier are separated from a neighbouring codon by a binding sequence. The binding region may have any suitable number of nucleotides, e.g. 1 to 20.

The binding region, if present, may serve various purposes besides serving as a substrate for an enzyme. In one setup of the invention, the binding region identifies the position of the codon. Usually, the binding region either upstream or downstream of a codon comprises information which allows determination of the position of the codon. In another setup, the binding regions have alternating sequences, allowing for addition of building blocks from two pools in the formation of the library. Moreover, the binding region may adjust the annealing temperature to a desired level.

A binding region with high affinity can be provided by incorporation of one or more nucleobases forming three hydrogen bonds to a cognate nucleobase. Examples of nucleobases having this property are guanine and cytosine. Alternatively, or in addition, the binding region may be subjected to backbone modification. Several backbone modifications provides for higher affinity, such as 2'-O-methyl substitution of the ribose moiety, peptide nucleic acids (PNA), and 2'-4' O-methylene cyclisation of the ribose moiety, also referred to as LNA (Locked Nucleic Acid).

The identifier may comprise flanking regions around the codons. The flanking region can encompass a signal group, such as a flourophor or a radio active group to allow for detection of the presence or absence of a complex or the flanking region may comprise a label that may be detected, such as biotin. When the identifier comprises a biotin moiety, the identifier may easily be recovered.

The flanking regions can also serve as priming sites for amplification reactions, such as PCR. Usually, the last cycle in the formation of the bifunctional complex includes the incorporation of a priming site. A region of the bifunctional complex close to the display molecule, such as a nucleic acid sequence between the display molecule and the codon coding for the scaffold molecule, is usually used for another priming site, thereby allowing for PCR amplification of the coding region of the bifunctional complex.

Combination of Mode 1 and Mode 2:

In a certain aspect of the invention, mode 1 and mode 2 described above is combined, i.e. different reactants are used in different rounds. Also within mode 1 and mode 2 different building blocks may be used in different rounds.

In the formation of a library it may be advantageous to use a combination of a one-pot synthesis strategy (mode 1) and a split-and-mix strategy (mode 2), because each of mode 1 and mode 2 has its virtues. The one-pot strategy offers the possibility of having the reactive groups in close proximity prior to reaction, thus obtaining a high local concentration and the convenience of having a single container. The split-and mix strategy offers the possibility of having a free reactant and non-hybridising reaction conditions, providing for versatile reactions. It may be appropriate to refer to Fig. 15 in which various single encoding enzymatic methods are shown. A split-and-mix synthesis strategy is generally used for reactants not having a covalent link between the reactant/functional entity and the codon/anti-codon, i.e. free reactants and zipper building blocks. A one-pot synthesis strategy is generally used for reactants in which a covalent link exist between the functional entity and the codon/anti-codon identifying said functional entity, i.e. the E2 building blocks, loop building blocks, and the N building blocks.

In a certain embodiment of the invention an intermediate library of bifunctional complexes is generated using a one-pot synthesis strategy. This intermediate library is subsequently used for the generation of a final library by a split-and-mix synthesis. The intermediate library may be generated using a single round or multiple rounds of one-pot synthesis and the final library may be produced applying a single or multiple rounds of split-and-mix. The use of a split-and-mix synthesis in the last round of library generation offers the possibility of using a reaction media not compatible with maintenance of a hybridisation, e.g. high ionic strength or organic solvents, for the final reactant.

In another embodiment an intermediate library is produced using a split and mix synthesis strategy. The intermediate library is used for the generation of a final library using a one-pot synthesis strategy. The intermediate library may be produced using a single or multiple rounds of split-and-mix synthesis and the final library may be manufactured applying one or more rounds of one-pot synthesis. The one-pot synthesis in the final round provide for a close proximity between the growing encoded molecule and the functional entity. The close proximity results in a high local concentration promoting the reaction even for reactants having a relatively low tendency to react.

Multiple encoding

Multiple encoding implies that two or more codons are provided in the identifier prior to or subsequent to a reaction between the chemical reactive site and two or more reactants. Multiple encoding has various advantages, such allowing a broader range of reactions possible, as many compounds can only be synthesis by a three (or more) component reaction because an intermediate between the first reactant and the chemical reactive site is not stable. Other advantages relates to the use of organic solvents and the availability of two or more free reactants in certain embodiments.

Thus in a certain aspect of the invention, it relates to a method for obtaining a bifunctional complex comprising a display molecule part and a coding part, wherein the display molecule is obtained by reaction of a chemical reactive site with two or more reactants and the coding part comprises tag(s) identifying the reactants.

In a certain aspect of the invention, a first reactant forms an intermediate product upon reaction with the chemical reactive site and a second reactant reacts with the intermediate product to obtain the display molecule or a precursor thereof. In another aspect of the invention, two or more reactants react with each other to form an intermediate product and the chemical reactive site reacts with this intermediate product to obtain the display molecule or a precursor thereof. The intermediate product can be obtained by reacting the two or more reactants separately and then in a subsequent step reacting the intermediate product with the chemical reactive site. Reacting the reactants in a separate step provide for the possibility of using conditions the tags would not withstand. Thus, in case the coding part comprises nucleic acids, the reaction between the reactant may be conducted at conditions that otherwise would degrade the nucleic acid.

The reactions may be carried out in accordance with the scheme shown below. The scheme shows an example in which the identifying tags for two reactants and the chemical reactive site (scaffold) attached to the chemical reaction site are provided in separate compartments. The compartments are arranged in an array, such as a micro-titer plate, allowing for any combination of the different acylating agents and the different alkylating agents.

Starting situation:

Alkylating agents / Acylating agents	A	B	C	...
1	Tagx11-X	Tagx12-X	Tagx13-X	...
2	Tagx21-X	Tagx22-X	Tagx23-X	...
3	Tagx31-X	Tagx32-X	Tagx33-X	...
...	...	...	...	...

X denotes a chemical reaction site such as a scaffold.

The two reactants are either separately reacted with each other in any combination or subsequently added to each compartment in accordance with the tags of the coding part or the reactants may be added in any order to each compartment to allow for a direct reaction. The scheme below shows the result of the reaction.

Plate of products

Alkylating agents / Acylating agents	A	B	C	...
1	Tagx11-XA1	Tagx12-XB1	Tagx13-XC1	...
2	Tagx21-XA2	Tagx22-XB2	Tagx23-XC2	...
3	Tagx31-XA3	Tagx32-XB3	Tagx33-XC3	...
...	...	...	...	...

As an example XA2 denotes display molecule XA2 in its final state, i.e. fully assembled from fragments X, A and 2.

The coding part comprising the two or more tags identifying the reactants, may be prepared in any suitable way either before or after the reaction. In one aspect of the invention, each of the coding parts are synthesised by standard phosphoramidite chemistry. In another aspect the tags are pre-prepared and assembled into the final coding part by chemical or enzymatic ligation.

Various possibilities for chemical ligation exist. Suitable examples include that
a) a first oligonucleotide end comprises a 3'-OH group and the second oligonucleotide end comprises a 5'-phosphor-2-imidazole group. When reacted a phosphodiester internucleoside linkage is formed,

b) a first oligonucleotide end comprising a phosphoimidazolide group and the 3'-end and a phosphoimidazolide group at the 5'-and. When reacted together a phosphodiester internucleoside linkage is formed,
c) a first oligonucleotide end comprising a 3'-phosphorothioate group and a second oligonucleotide comprising a 5'-iodine. When the two groups are reacted a 3'-O-P(=O)(OH)-S-5' internucleoside linkage is formed, and
d) a first oligonucleotide end comprising a 3'-phosphorothioate group and a second oligonucleotide comprising a 5'-tosylate. When reacted a 3'-O-P(=O)(OH)-S-5' internucleoside linkage is formed.

Suitably, the tags operatively are joined together, so that as to allow a nucleic acid active enzyme to recognize the ligation area as substrate. Notably, in a preferred embodiment, the ligation is performed so as to allow a polymerase to recognise the ligated strand as a template. Thus, in a preferred aspect, a chemical reaction strategy for the coupling step generally includes the formation of a phosphodiester internucleoside linkage. In accordance with this aspect, method a) and b) above are preferred.

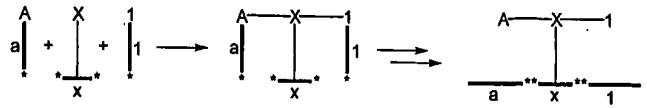
In another aspect, when ligases are used for the ligation, suitable examples include Taq DNA ligase, T4 DNA ligase, T7 DNA ligase, and *E. coli* DNA ligase. The choice of the ligase depends to a certain degree on the design of the ends to be joined together. Thus, if the ends are blunt, T4 DNA ligase may be preferred, while a Taq DNA ligase may be preferred for a sticky end ligation, i.e. a ligation in which an overhang on each end is a complement to each other.

In a certain aspect of the invention enzymatic encoding is preferred because of the specificity enzymes provide. Fig. 17 discloses a variety of methods for enzymatically encoding two or more reactants in the coding part of the bifunctional molecule. The choice of encoding method depends on a variety of factors, such as the need for free reactants, the need for proximity, and the need for convenience. The enzymatic double encoding methods shown on Fig. 17 may easily be expanded to triple, quarto, etc. encoding.

In accordance with a certain embodiment functional entities are attached to identifying tags, and each functional entity carries one or more reactive groups. All the functional entities react with each other to generate the final product containing as many tags as

functional entities. The tags may be combined into a single coding part, usually an oligonucleotide through an intermolecular reaction or association followed by cleavage of two of the linkers, as shown below:

A.

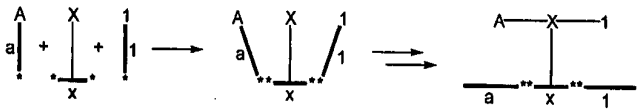


Bold lines represent tags. Thin lines represent linkers or bonds. "*" denotes a priming site. In some aspects of the invention X is regarded as the chemical reactive site.

In one aspect of the above embodiment the tags are of oligonucleotides, which combine through chemical ligation or enzyme catalysed ligation.

Alternatively, the tags are coupled prior to the reaction of the functional entities. In that process the functional entities will be cleaved from their tags or cleaved afterwards. E.g.

B.



An embodiment of the above schematic representation comprises, when the tags are nucleotides, the combination of tags through chemical ligation or enzyme catalysed ligation.

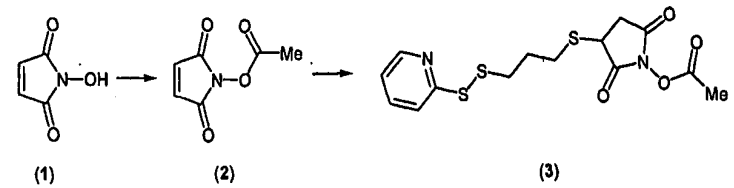
Example 9 illustrates a multi component reaction in which triple encoding is used. Thus after the reaction of three free reactants with a chemical reactive site, the coding part is provided with three identifying tags by enzymatic ligation.

Building blocks capable of transferring functional entities.

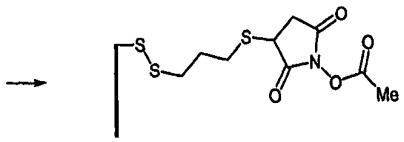
The following sections describe the formation and use of exemplary building blocks capable of transferring a functional entity to a reactive group of a bifunctional complex. A bold line indicates an oligonucleotide.

A. Acylation reactions

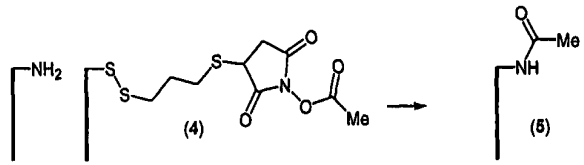
General route to the formation of acylating building blocks and the use of these:



(1) (2) (3)



(4)

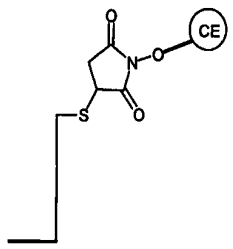


N-hydroxymaleimide (1) may be acylated by the use of an acylchloride e.g. acetylchloride or alternatively acylated in e.g. THF by the use of dicyclohexylcarbodiimide or diisopropylcarbodiimide and acid e.g. acetic acid. The intermediate may be subjected to Michael addition by the use of excess 1,3-propanedithiol, followed by reaction with either 4,4'-dipyridyl disulfide or 2,2'-dipyridyl disulfide. This intermediate (3) may then be loaded onto an oligonucleotide carrying a thiol handle to generate the building block (4). Obviously, the intermediate (2) can be attached to the oligonucleotide using another linkage than the disulfide linkage, such as an amide linkage and the *N*-hydroxymaleimide can be distanced from the oligonucleotide using a variety of spacers.

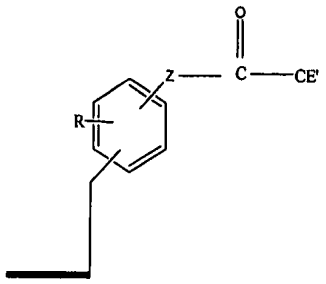
The building block (4) may be reacted with an identifier oligonucleotide comprising a recipient amine group e.g. by following the procedure: The building block (4) (1 nmol) is

mixed with an amino-oligonucleotide (1 nmol) in hepes-buffer (20 µL of a 100 mM hepes and 1 M NaCl solution, pH=7.5) and water (39 uL). The oligonucleotides are annealed together by heating to 50 °C and cooling (2 °C/ second) to 30 °C. The mixture is then left o/n at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second), to yield the product (5).

In more general terms, the building blocks indicated below is capable of transferring a chemical entity (CE) to a recipient nucleophilic group, typically an amine group. The bold lower horizontal line illustrates the building block and the vertical line illustrates a spacer. The 5-membered substituted N-hydroxysuccinimid (NHS) ring serves as an activator, i.e. a labile bond is formed between the oxygen atom connected to the NHS ring and the chemical entity. The labile bond may be cleaved by a nucleophilic group, e.g. positioned on a scaffold



Another building block which may form an amide bond is



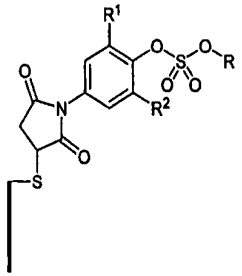
R may be absent or NO_2, CF_3, halogen, preferably Cl, Br, or I, and Z may be S or O. This type of building block is disclosed in Danish patent application No. PA 2002 0951 and US provisional patent application filed 20 December 2002 with the title "A building block capable of transferring a functional entity to a recipient reactive group". The content of both patent application are incorporated herein in their entirety by reference..

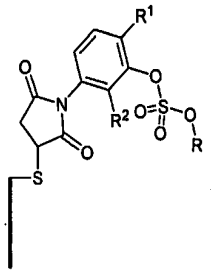
A nucleophilic group can cleave the linkage between Z and the carbonyl group thereby transferring the chemical entity –(C=O)-CE' to said nucleophilic group.

B. Alkylation

General route to the formation of alkylating/vinylating building blocks and use of these:

Alkylating building blocks may have the following general structure:

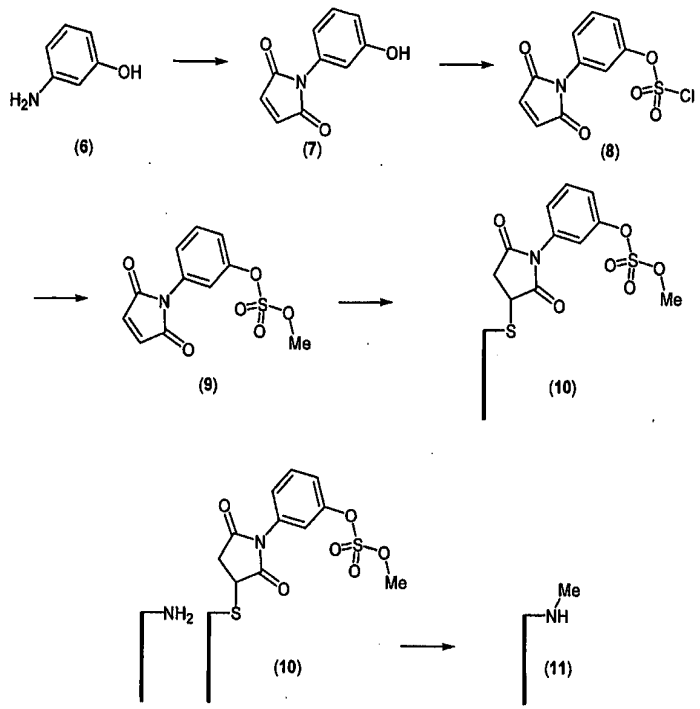




R^1 = H, Me, Et, iPr, Cl, NO_2
R^2 = H, Me, Et, iPr, Cl, NO_2

R^1 and R^2 may be used to tune the reactivity of the sulphate to allow appropriate reactivity. Chloro and nitro substitution will increase reactivity. Alkyl groups will decrease reactivity. Ortho substituents to the sulphate will due to steric reasons direct incoming nucleophiles to attack the R-group selectively and avoid attack on sulphur.

An example of the formation of an alkylating building block and the transfer of a functional entity is depicted below:



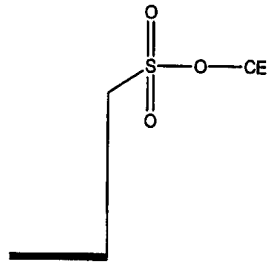
3-Aminophenol (**6**) is treated with maleic anhydride, followed by treatment with an acid e.g. H_2SO_4 or P_2O_5 and heated to yield the maleimide (**7**). The ring closure to the maleimide may also be achieved when an acid stable *O*-protection group is used by treatment with Ac_2O, with or without heating, followed by *O*-deprotection. Alternatively reflux in Ac_2O, followed by *O*-deacetylation in hot water/dioxane to yield (**7**). Further treatment of (**7**) with SO_2Cl_2, with or without triethylamine or potassium carbonate in dichloromethane or a higher boiling solvent will yield the intermediate (**8**), which may be isolated or directly further transformed into the aryl alkyl sulphate by the quench with the appropriate alcohol, in this case MeOH, whereby (**9**) will be formed.

The organic moiety (**9**) may be connected to an oligonucleotide, as follows: A thiol carrying oligonucleotide in buffer 50 mM MOPS or hepes or phosphate pH 7.5 is treated with a 1-100 mM solution and preferably 7.5 mM solution of the organic building block (**9**) in DMSO or alternatively DMF, such that the DMSO/DMF concentration is 5-50%,

and preferably 10%. The mixture is left for 1-16 h and preferably 2-4 h at 25 °C to give the alkylating agent in this case a methylating building block (**10**).

The reaction of the alkylating building block (**10**) with an amine bearing nascent bifunctional complex may be conducted as follows: The bifunctional complex (1 nmol) is mixed the building block (**10**) (1 nmol) in hepes-buffer (20 µL of a 100 mM hepes and 1 M NaCl solution, pH=7.5) and water (39 uL). The oligonucleotides are annealed to each other by heating to 50 °C and cooled (2 °C/ second) to 30 °C. The mixture is then left o/n at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second), to yield the methylamine reaction product (**11**).

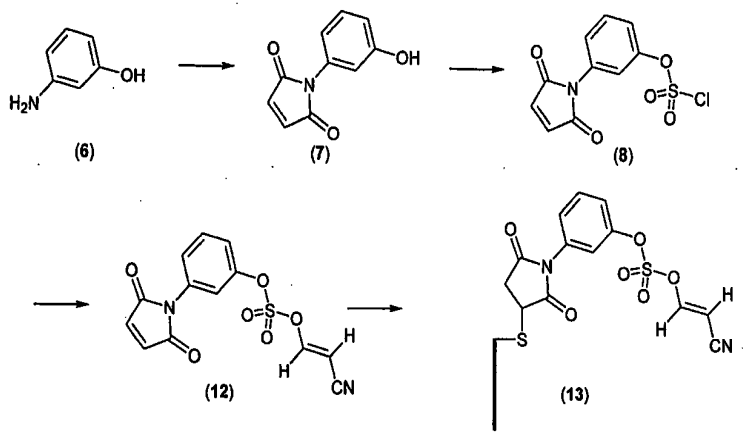
In more general terms, a building block capable of transferring a chemical entity to a receiving reactive group forming a single bond is



The receiving group may be a nucleophile, such as a group comprising a hetero atom, thereby forming a single bond between the chemical entity and the hetero atom, or the receiving group may be an electronegative carbon atom, thereby forming a C-C bond between the chemical entity and the scaffold.

C. Vinylation reactions

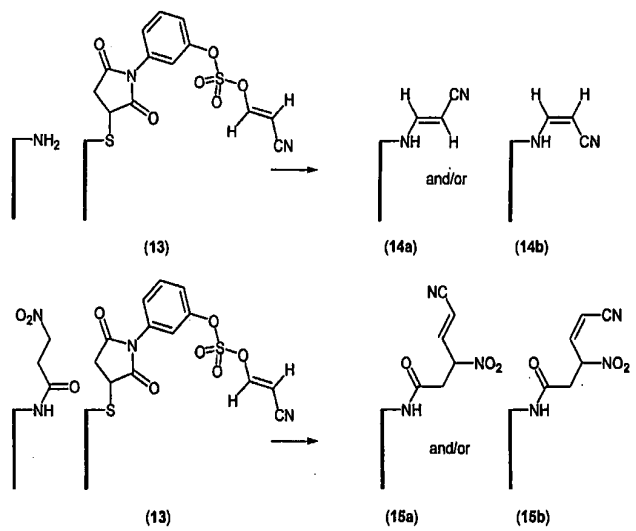
A vinylating building block may be prepared and used similarly as described above for an alkylating building block. Although instead of reacting the chlorosulphonate (**8** above) with an alcohol, the intermediate chlorosulphate is isolated and treated with an enolate or *O*-trialkylsilylenolate with or without the presence of fluoride. E.g.



Formation of an exemplary vinylating building block (**13**):

The thiol carrying oligonucleotide in buffer 50 mM MOPS or hepes or phosphate pH 7.5 is treated with a 1-100 mM solution and preferably 7.5 mM solution of the organic moiety (**12**) in DMSO or alternatively DMF, such that the DMSO/DMF concentration is 5-50%, and preferably 10%. The mixture is left for 1-16 h and preferably 2-4 h at 25 °C to give the vinylating building block (**13**).

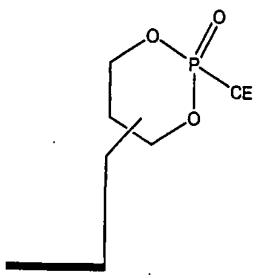
The sulfonylenolate (**13**) may be used to react with amine carrying scaffold to give an enamine (**14a** and/or **14b**) or e.g. react with a carbanion to yield (**15a** and/or **15b**). E.g.



The reaction of the vinylating building block (**13**) and an amine or nitroalkyl carrying identifier may be conducted as follows:

The amino-oligonucleotide (1 nmol) or nitroalkyl-oligonucleotide (1 nmol) identifier is mixed with the building block (1 nmol) (**13**) in 0.1 M TAPS, phosphate or hepes-buffer and 300 mM NaCl solution, pH=7.5-8.5 and preferably pH=8.5. The oligonucleotides are annealed to the template by heating to 50 °C and cooled (2 °C/ second) to 30 °C. The mixture is then left o/n at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second), to yield reaction product (**14a/b** or **15a/b**). Alternative to the alkyl and vinyl sulphates described above may equally effective be sulphonates as e.g. (**31**) (however with R" instead as alkyl or vinyl), described below, prepared from (**28**, with the phenyl group substituted by an alkyl group) and (**29**), and be used as alkylating and vinylating agents.

Another building block capable of forming a double bond by the transfer of a chemical entity to a recipient aldehyde group is shown below. A double bond between the carbon of the aldehyde and the chemical entity is formed by the reaction.



The above building block is comprised by the Danish patent application No. DK PA 2002 01952 and the US provisional patent application filed 20 December 2002 with the title "A building block capable of transferring a functional entity to a recipient reactive group forming a C=C double bond". The content of both patent applications are incorporated herein in their entirety by reference.

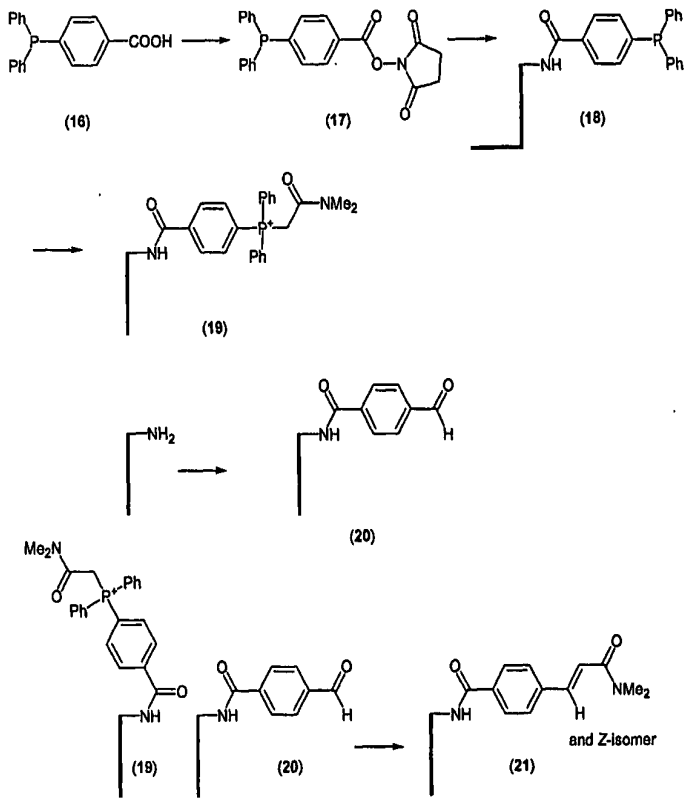
D. Alkenylidation reactions

General route to the formation of Wittig and HWE building blocks and use of these:

Commercially available compound (**16**) may be transformed into the NHS ester (**17**) by standard means, *i.e.* DCC or DIC couplings. An amine carrying oligonucleotide in buffer 50 mM MOPS or hepes or phosphate pH 7.5 is treated with a 1-100 mM solution and preferably 7.5 mM solution of the organic compound in DMSO or alternatively DMF, such that the DMSO/DMF concentration is 5-50%, and preferably 10%. The mixture is left for 1-16 h and preferably 2-4 h at 25 °C to give the phosphine bound precursor building block (**18**). This precursor building block is further transformed by addition of the appropriate alkylhalide, e.g. *N,N*-dimethyl-2-iodoacetamide as a 1-100 mM and preferably 7.5 mM solution in DMSO or DMF such that the DMSO/DMF concentration is 5-50%, and preferably 10%. The mixture is left for 1-16 h and preferably 2-4 h at 25 °C to give the building block (**19**). As an alternative to this, the organic compound (**17**) may be *P*-alkylated with an alkylhalide and then be coupled onto an amine carrying oligonucleotide to yield (**19**).

An aldehyde carrying identifier (**20**), may be formed by the reaction between the NHS ester of 4-formylbenzoic acid and an amine carrying oligonucleotide, using conditions

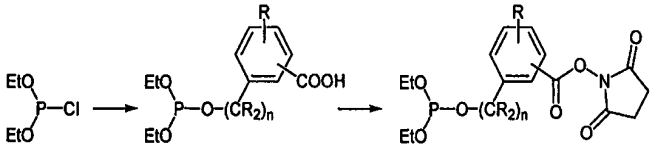
similar to those described above. The identifier (**20**) reacts with (**19**) under slightly alkaline conditions to yield the alkene (**21**).



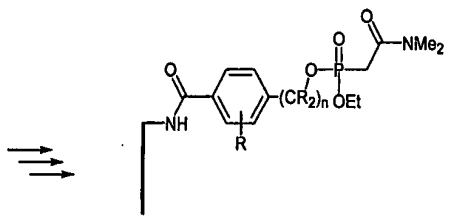
The reaction of monomer building blocks (**19**) and identifier (**20**) may be conducted as follows: The identifier (**20**) (1 nmol) is mixed with building block (**19**) (1 nmol) in 0.1 M TAPS, phosphate or hepes-buffer and 1 M NaCl solution, pH=7.5-8.5 and preferably pH=8.0. The reaction mixture is left at 35-65 °C preferably 58 °C over night to yield reaction product (**21**).

As an alternative to (**17**), phosphonates (**24**) may be used instead. They may be prepared by the reaction between diethylchlorophosphite (**22**) and the appropriate carboxy

carrying alcohol. The carboxylic acid is then transformed into the NHS ester (**24**) and the process and alternatives described above may be applied. Although instead of a simple *P*-alkylation, the phosphite may undergo Arbuzov's reaction and generate the phosphonate. Building block (**25**) benefits from the fact that it is more reactive than its phosphonium counterpart (**19**).



(**22**) (**23**) n=0-2 (**24**)



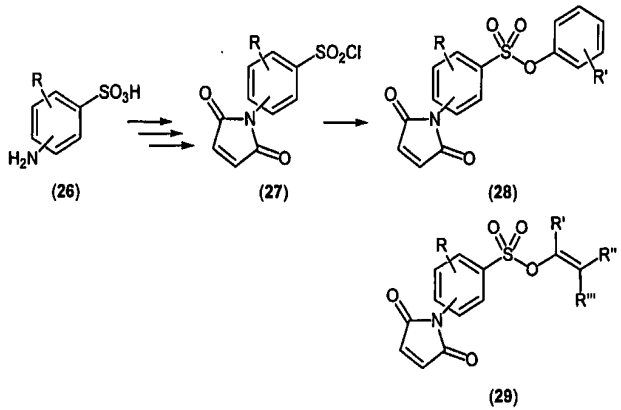
(**25**)

E. Transition metal catalyzed arylation, hetarylation and vinylation reactions

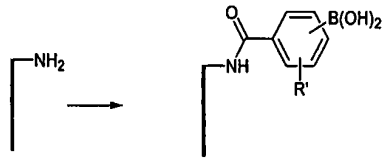
Electrophilic building blocks (**31**) capable of transferring an aryl, hetaryl or vinyl functionality may be prepared from organic compounds (**28**) and (**29**) by the use of coupling procedures for maleimide derivatives to SH-carrying oligonucleotides described above. Alternatively to the maleimide the NHS-ester derivatives may be prepared from e.g. carboxybenzensulfonic acid derivatives, be used by coupling of these to an amine carrying oligonucleotide. The R-group of (**28**) and (**29**) is used to tune the reactivity of the sulphonate to yield the appropriate reactivity.

The transition metal catalyzed cross coupling may be conducted as follows: A premix of 1.4 mM Na_2PdCl_4 and 2.8 mM $P(p\text{-}SO_3C_6H_4)_3$ in water left for 15 min was added to a mixture of the identifier (**30**) and building block (**31**) (both 1 nmol) in 0.5 M NaOAc

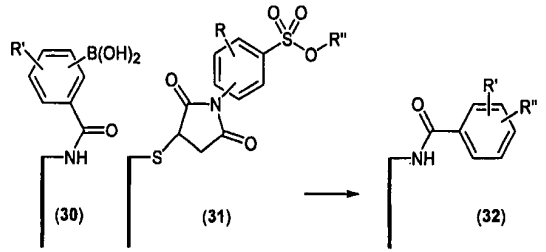
buffer at pH=5 and 75 mM NaCl (final [Pd]=0.3 mM). The mixture is then left o/n at 35-65 °C preferably 58 °C, to yield reaction product (**32**).



(**26**) (**27**) (**28**) (**29**)



(**30**)



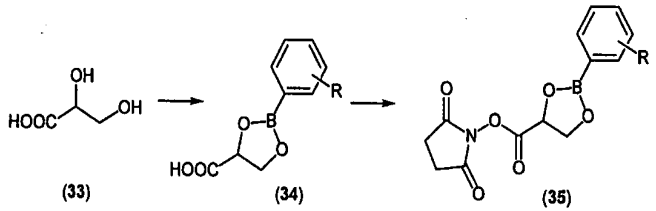
(**30**) (**31**) (**32**)

R" = aryl, hetaryl or vinyl

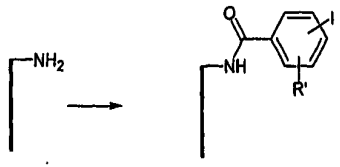
Corresponding nucleophilic monomer building blocks capable of transferring an aryl, hetaryl or vinyl functionality may be prepared from organic compounds of the type (**35**).

This is available by estrification of a boronic acid by a diol e.g. (33), followed by transformation into the NHS-ester derivative. The NHS-ester derivative may then be coupled to an oligonucleotide, by use of coupling procedures for NHS-ester derivatives to amine carrying oligonucleotides described above, to generate building block type (37). Alternatively, maleimide derivatives may be prepared as described above and loaded onto SH-carrying oligonucleotides.

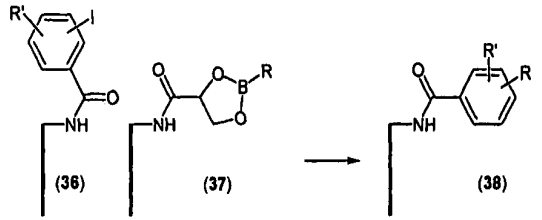
The transition metal catalyzed cross coupling is conducted as follows:
A premix of 1.4 mM Na_2PdCl_4 and 2.8 mM $P(p\text{-}SO_3C_6H_4)_3$ in water left for 15 min was added to a mixture of the identifier (36) and the building block (37) (both 1 nmol) in 0.5 M NaOAc buffer at pH=5 and 75 mM NaCl (final [Pd]=0.3 mM). The mixture is then left o/n at 35-65 °C preferably 58 °C, to yield template bound (38).



(33) (34) (35)



(36)

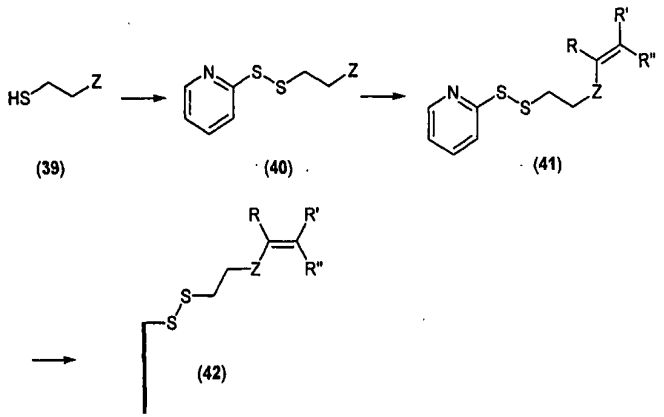


(36) (37) (38)

R = aryl, hetaryl or vinyl

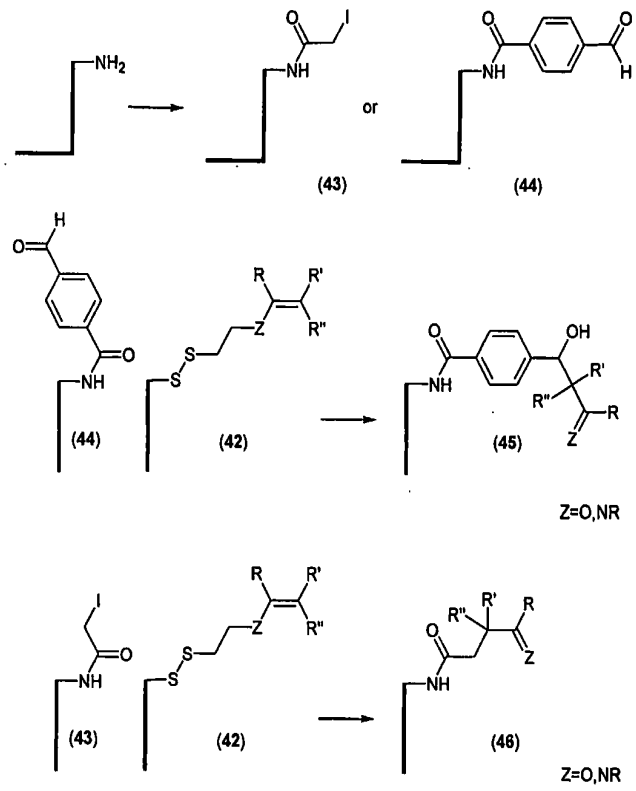
F. Reactions of enamine and enolether monomer building blocks

Building blocks loaded with enamines and enolethers may be prepared as follows:
For Z=NHR (R=H, alkyl, aryl, hetaryl), a 2-mercaptoethylamine may be reacted with a dipyridyl disulfide to generate the activated disulfide (40), which may then be condensed to a ketone or an aldehyde under dehydrating conditions to yield the enamine (41). For Z=OH, 2-mercaptoethanol is reacted with a dipyridyl disulfide, followed by *O*-tosylation (Z=OTs). The tosylate (40) may then be reacted directly with an enolate or in the presence of fluoride with a O-trialkylsilylenolate to generate the enolate (41).
The enamine or enolate (41) may then be coupled onto an SH-carrying oligonucleotide as described above to give the building block (42).

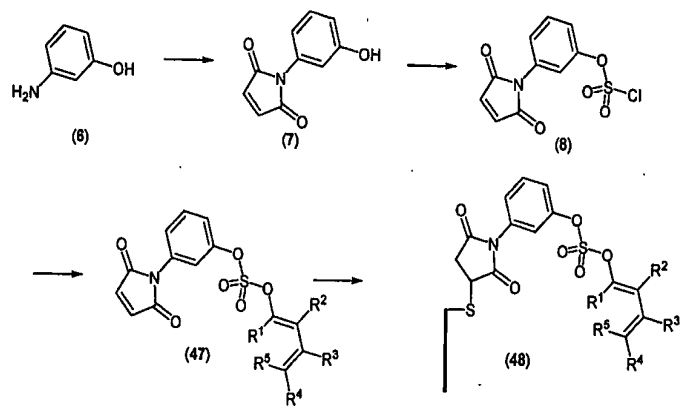


The building block (**42**) may be reacted with a carbonyl carrying identifier oligonucleotide like (**44**) or alternatively an alkylhalide carrying oligonucleotide like (**43**) as follows: The building block (**42**) (1 nmol) is mixed with the identifier (**43**) (1 nmol) in 50 mM MOPS, phosphate or hepes-buffer buffer and 250 mM NaCl solution, pH=7.5-8.5 and preferably pH=7.5. The reaction mixture is left at 35-65 °C preferably 58 °C over night or alternatively at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second) to yield reaction product (**46**), where Z=O or NR. For compounds where Z=NR slightly acidic conditions may be applied to yield product (**46**) with Z=O.

The building block (**42**) (1 nmol) is mixed with the identifier (**44**) (1 nmol) in 0.1 M TAPS, phosphate or hepes-buffer buffer and 300 mM NaCl solution, pH=7.5-8.5 and preferably pH=8.0. The reaction mixture is left at 35-65 °C preferably 58 °C over night or alternatively at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second) to yield reaction product (**45**), where Z=O or NR. For compounds where Z=NR slightly acidic conditions may be applied to yield product (**45**) with Z=O.

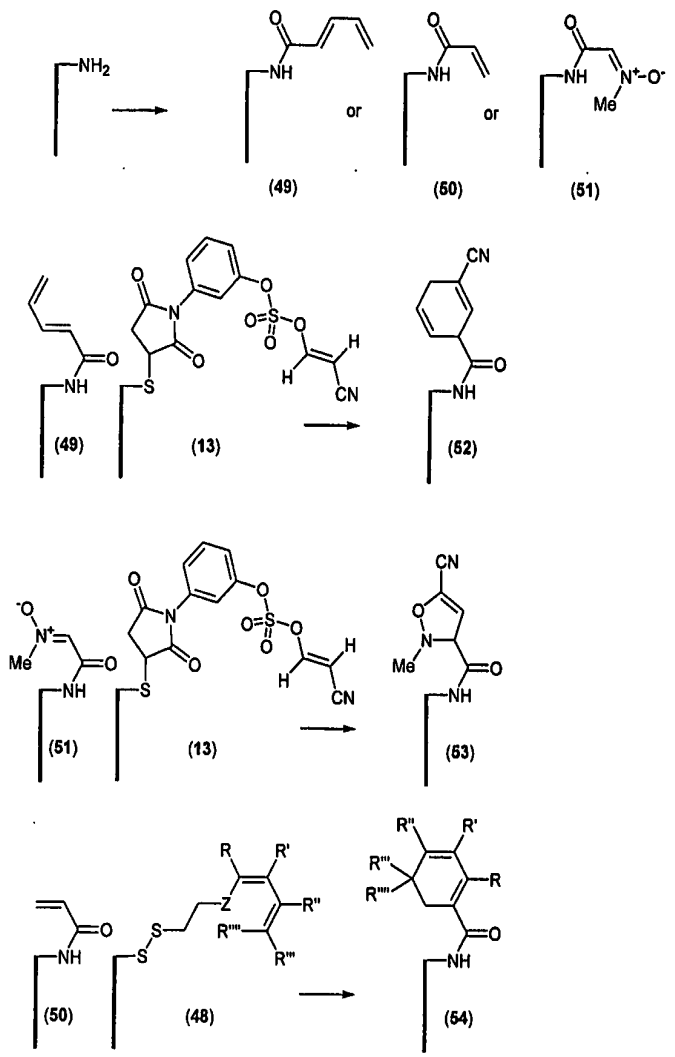


Enolethers type (**13**) may undergo cycloaddition with or without catalysis. Similarly, dienolethers may be prepared and used, e.g. by reaction of (**8**) with the enolate or trialkylsilylenolate (in the presence of fluoride) of an α,β-unsaturated ketone or aldehyde to generate (**47**), which may be loaded onto an SH-carrying oligonucleotide, to yield monomer building block (**48**).



The diene (**49**), the ene (**50**) and the 1,3-dipole (**51**) may be formed by simple reaction between an amino carrying oligonucleotide and the NHS-ester of the corresponding organic compound. Reaction of (**13**) or alternatively (**31**, R"=vinyl) with dienes as e.g. (**49**) to yield (**52**) or e.g. 1,3-dipoles (**51**) to yield (**53**) and reaction of (**48**) or (**31**, R"=dienyl) with enes as e.g. (**50**) to yield (**54**) may be conducted as follows:

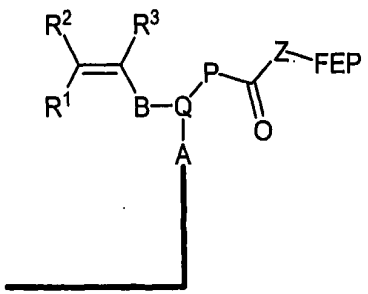
The building block (**13**) or (**48**) (1 nmol) is mixed with the identifier (**49**) or (**50**) or (**51**) (1 nmol) in 50 mM MOPS, phosphate or hepes-buffer buffer and 2.8 M NaCl solution, pH=7.5-8.5 and preferably pH=7.5. The reaction mixture is left at 35-65 °C preferably 58 °C over night or alternatively at a fluctuating temperature (10 °C for 1 second then 35 °C for 1 second) to yield template bound (**52**), (**53**) or (**54**), respectively.



Cross-link cleavage building blocks

It may be advantageous to split the transfer of a chemical entity to a recipient reactive group into two separate steps, namely a cross-linking step and a cleavage step be-

cause each step can be optimized. A suitable building block for this two step process is illustrated below:



Initially, a reactive group appearing on the functional entity precursor (abbreviated FEP) reacts with a recipient reactive group, e.g. a reactive group appearing on a scaffold, thereby forming a cross-link. Subsequently, a cleavage is performed, usually by adding an aqueous oxidising agent such as I_2, Br_2, Cl_2, H^+, or a Lewis acid. The cleavage results in a transfer of the group HZ-FEP- to the recipient moiety, such as a scaffold.

In the above formula

Z is O, S, NR^4

Q is N, CR^1

P is a valence bond, O, S, NR^4, or a group C_{5-7}arylene, C_{1-6}alkylene, C_{1-6}O-alkylene, C_{1-6}S-alkylene, NR^1-alkylene, C_{1-6}alkylene-O, C_{1-6}alkylene-S option said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

B is a group comprising D-E-F, in which

D is a valence bond or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

E is, when present, a valence bond, O, S, NR^4, or a group C_{1-6}alkylene, C_{1-6}alkenylene, C_{1-6}alkynylene, C_{5-7}arylene, or C_{5-7}heteroarylene, said group optionally being substituted with 1 to 4 group R^{11},

F is, when present, a valence bond, O, S, or NR^4,

A is a spacing group distancing the chemical structure from the complementing element, which may be a nucleic acid,

R^1, R^2, and R^3 are independent of each other selected among the group consisting of H, C_1-C_8 alkyl, C_2-C_8 alkenyl, C_2-C_8 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

FEP is a group selected among the group consisting of H, C_1-C_8 alkyl, C_2-C_8 alkenyl, C_2-C_8 alkynyl, C_4-C_8 alkadienyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, and heteroaryl, said group being substituted with 0-3 R^4, 0-3 R^5 and 0-3 R^9 or C_1-C_3 alkylene-NR^4_2, C_1-C_3 alkylene-$NR^4C(O)R^8$, C_1-C_3 alkylene-$NR^4C(O)OR^8$, C_1-C_2 alkylene-O-NR^4_2, C_1-C_2 alkylene-O-$NR^4C(O)R^8$, C_1-C_2 alkylene-O-$NR^4C(O)OR^8$ substituted with 0-3 R^9,

where R^4 is H or selected independently among the group consisting of C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, C_3-C_7 cycloheteroalkyl, aryl, heteroaryl, said group being substituted with 0-3 R^9 and

R^5 is selected independently from $-N_3$, -CNO, $-C(NOH)NH_2$, -NHOH, $-NHNHR^6$, $-C(O)R^6$, $-SnR^6_3$, $-B(OR^6)_2$, $-P(O)(OR^6)_2$ or the group consisting of C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_4-C_8 alkadienyl said group being substituted with 0-2 R^7,

where R^6 is selected independently from H, C_1-C_6 alkyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-5 halogen atoms selected from -F, -Cl, -Br, and -I; and R^7 is independently selected from $-NO_2$, $-COOR^6$, $-COR^6$, -CN, $-OSiR^6_3$, $-OR^6$ and $-NR^6_2$.

R^8 is H, C_1-C_6 alkyl, C_2-C_6 alkenyl, C_2-C_6 alkynyl, C_3-C_7 cycloalkyl, aryl or C_1-C_6 alkylene-aryl substituted with 0-3 substituents independently selected from -F, -Cl, $-NO_2$, $-R^3$, $-OR^3$, $-SiR^3_3$

R^9 is =O, -F, -Cl, -Br, -I, -CN, $-NO_2$, $-OR^6$, $-NR^6_2$, $-NR^6$-$C(O)R^6$, $-NR^6$-$C(O)OR^6$, $-SR^6$, $-S(O)R^6$, $-S(O)_2R^6$, $-COOR^6$, $-C(O)NR^6_2$ and $-S(O)_2NR^6_2$.

In a preferred embodiment Z is O or S, P is a valence bond, Q is CH, B is CH_2, and R^1, R^2, and R^3 is H. The bond between the carbonyl group and Z is cleavable with aqueous I_2.

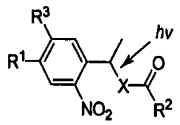
Cleavable linkers

A cleavable linker may be positioned between the target and a solid support or between the potential drug candidate and the identifier region or any other position that may ensure a separation of the nucleic acid sequence comprising the codons from successful complexes from non-specific binding complexes. The cleavable linker may be selectively cleavable, i.e. conditions may be selected that only cleave that particular linker.

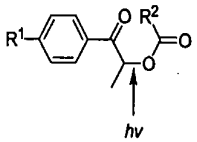
The cleavable linkers may be selected from a large plethora of chemical structures. Examples of linkers includes, but are not limited to, linkers having an enzymatic cleavage site, linkers comprising a chemical degradable component, linkers cleavable by electromagnetic radiation.

Examples of linkers cleavable by electromagnetic radiation (light)

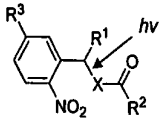
o-nitrobenzyl



p-alkoxy

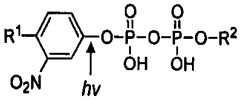


o-nitrobenzyl in exo position



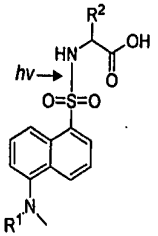
For more details see Holmes CP. J. Org. Chem. **1997**, *62*, 2370-2380

3-nitrophenyloxy



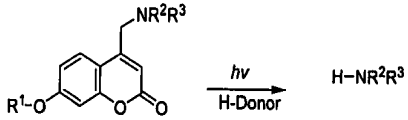
For more details see Rajasekharan Pillai, V. N. Synthesis. **1980**, 1-26

Dansyl derivatives:



For more details see Rajasekharan Pillai, V. N. Synthesis. **1980**, 1-26

Coumarin derivatives



For more details see R. O. Schoenleber, B. Giese. Synlett **2003**, 501-504

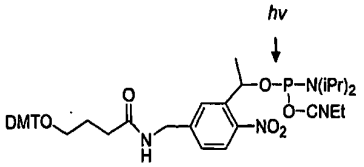
R^1 and R^2 can be either of the potential drug candidate and the identifier, respectively. Alternatively, R^1 and R^2 can be either of the target or a solid support, respectively.

R^3 = H or OCH_3

If X is O then the product will be a carboxylic acid

If X is NH the product will be a carboxamide

One specific example is the PC Spacer Phosphoramidite (Glen research catalog # 10-4913-90) which can be introduced in an oligonucleotide during synthesis and cleaved by subjecting the sample in water to UV light (~ 300-350 nm) for 30 seconds to 1 minute.

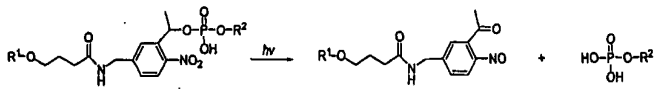


DMT = 4,4'-Dimethoxytrityl

iPr = Isopropyl

CNEt = Cyanoethyl

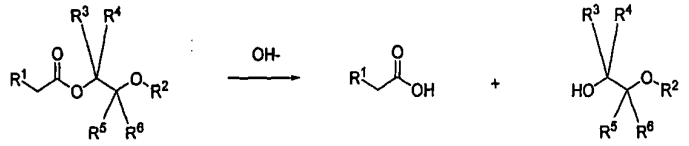
The above PC spacer phosphoamidite is suitable incorporated in a library of complexes at a position between the indentifier and the potential drug candidate. The spacer may be cleaved according to the following reaction.



R^1 and R^2 can be either of the encoded molecule and the identifying molecule, respectively. In a preferred aspect R^2 is an oligonucleotide identifier and the R^1 is the potential drug candidate. When the linker is cleaved a phosphate group is generated allowing for further biological reactions. As an example, the phosphate group may be positioned in the 5'end of an oligonucleotide allowing for an enzymatic ligation process to take place.

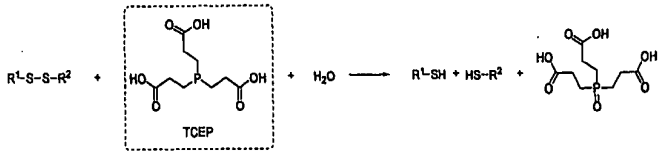
Examples of linkers cleavable by chemical agents:

Ester linkers can be cleaved by nucleophilic attack using e.g. hydroxide ions. In practice this can be accomplished by subjecting the target-ligand complex to a base for a short period.



R^1 and R^2 can be the either of be the potential drug candidate or the identifier, respectively. R^{4-6} can be any of the following: H, CN, F, NO_2, SO_2NR_2.

Disulfide linkers can efficiently be cleaved / reduced by Tris (2-carboxyethyl) phosphine (TCEP). TCEP selectively and completely reduces even the most stable water-soluble alkyl disulfides over a wide pH range. These reductions frequently required less than 5 minutes at room temperature. TCEP is a non-volatile and odorless reductant and unlike most other reducing agents, it is resistant to air oxidation. Trialkylphosphines such as TCEP are stable in aqueous solution, selectively reduce disulfide bonds, and are essentially unreactive toward other functional groups commonly found in proteins.



More details on the reduction of disulfide bonds can be found in Kirley, T.L.(1989), Reduction and fluorescent labeling of cyst(e)ine-containing proteins for subsequent structural analysis, *Anal. Biochem.* **180**, 231 and Levison, M.E., *et al.* (1969), Reduction of biological substances by water-soluble phosphines: Gamma-globulin. *Experentia* **25**, 126-127.

Linkers cleavable by enzymes

The linker connecting the potential drug candidate with the identifier or the solid support and the target can include a peptide region that allows a specific cleavage using a protease. This is a well-known strategy in molecular biology. Site-specific proteases and their cognate target amino acid sequences are often used to remove the fusion protein tags that facilitate enhanced expression, solubility, secretion or purification of the fusion protein.

Various proteases can be used to accomplish a specific cleavage. The specificity is especially important when the cleavage site is presented together with other sequences such as for example the fusion proteins. Various conditions have been optimized in order to enhance the cleavage efficiency and control the specificity. These conditions are available and know in the art.

Enterokinase is one example of an enzyme (serine protease) that cut a specific amino acid sequence. Enterokinase recognition site is Asp-Asp-Asp-Asp-Lys (DDDDK), and it cleaves C-terminally of Lys. Purified recombinant Enterokinase is commercially available and is highly active over wide ranges in pH (pH 4.5-9.5) and temperature (4-45°C).

The nuclear inclusion protease from tobacco etch virus (TEV) is another commercially available and well-characterized proteases that can be used to cut at a specific amino acid sequence. TEV protease cleaves the sequence Glu-Asn-Leu-Tyr-Phe-Gln-Gly/Ser (ENLYFQG/S) between Gln-Gly or Gln-Ser with high specificity.

Another well-known protease is thrombin that specifically cleaves the sequence Leu-Val-Pro-Arg-Gly-Ser (LVPAGS) between Arg-Gly. Thrombin has also been used for cleavage of recombinant fusion proteins. Other sequences can also be used for thrombin cleavage; these sequences are more or less specific and more or less efficiently cleaved by thrombin. Thrombin is a highly active protease and various reaction conditions are known to the public.

Activated coagulation factor FX (FXa) is also known to be a specific and useful protease. This enzyme cleaves C-terminal of Arg at the sequence Ile-Glu-Gly-Arg (IEGR). FXa is frequently used to cut between fusion proteins when producing proteins with recombinant technology. Other recognition sequences can also be used for FXa.

Other types of proteolytic enzymes can also be used that recognize specific amino acid sequences. In addition, proteolytic enzymes that cleave amino acid sequences in an un-specific manner can also be used if only the linker contains an amino acid sequence in the complex molecule.

Other type of molecules such as ribozymes, catalytically active antibodies, or lipases can also be used. The only prerequisite is that the catalytically active molecule can cleave the specific structure used as the linker, or as a part of the linker, that connects the encoding region and the displayed molecule or, in the alternative the solid support and the target.

A variety of endonucleases are available that recognize and cleave a double stranded nucleic acid having a specific sequence of nucleotides. The endonuclease Eco RI is an example of a nuclease that efficiently cuts a nucleotide sequence linker comprising the sequence GAATTC also when this sequence is close to the nucleotide sequence length. Purified recombinant Eco RI is commercially available and is highly active in a range of buffer conditions. As an example the Eco RI is working in in various protocols as indicted below (NEBuffer is available from New England Biolabs):

NEBuffer 1 : [10 mM Bis Tris Propane-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (pH 7.0 at 25°C)],

NEBuffer 2 : [50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (pH 7.9 at 25°C)],

NEBuffer 3 : [100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (pH 7.9 at 25°C)],

NEBuffer 4 : [50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM dithiothreitol (pH 7.9 at 25°C)].

Extension buffer : mM KCl, 20 mM Tris-HCl(Ph 8.8 at 25° C), 10 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$ and 0.1% Triton X-100, and 200 µM dNTPs.

Nucleotides

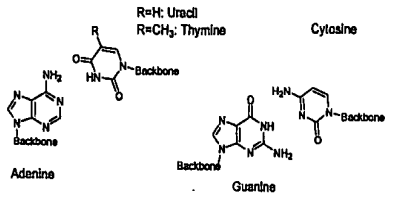
The nucleotides used in the present invention may be linked together in a sequence of nucleotides, i.e. an oligonucleotide. Each nucleotide monomer is normally composed of two parts, namely a nucleobase moiety, and a backbone. The back bone may in some cases be subdivided into a sugar moiety and an internucleoside linker.

The nucleobase moiety may be selected among naturally occurring nucleobases as well as non-naturally occurring nucleobases. Thus, "nucleobase" includes not only the known purine and pyrimidine hetero-cycles, but also heterocyclic analogues and tautomers thereof. Illustrative examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, purine, xanthine, diaminopurine, 8-oxo-N^6-methyladenine, 7-deazaxanthine, 7-deazaguanine, N^4,N^4-ethanocytosin, N^6,N^6-ethano-2,6-diaminopurine, 5-methylcytosine, 5-(C^3-C^6)-alkynylcytosine, 5-fluorouracil, 5-bromouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine and the "non-naturally occurring" nucleobases described in Benner et al., U.S. Pat No. 5,432,272. The term "nucleobase" is intended to cover these examples as well as analogues and tautomers thereof. Especially interesting nucleobases are adenine,

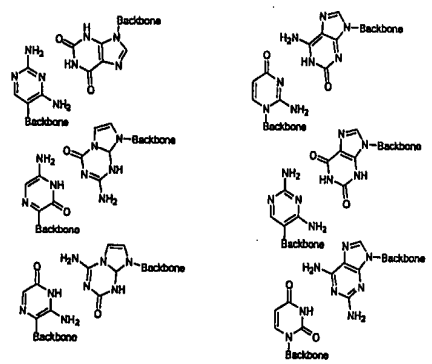
guanine, thymine, cytosine, 5-methylcytosine, and uracil, which are considered as the naturally occurring nucleobases.

Examples of suitable specific pairs of nucleobases are shown below:

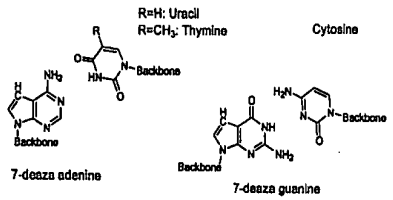
Natural Base Pairs



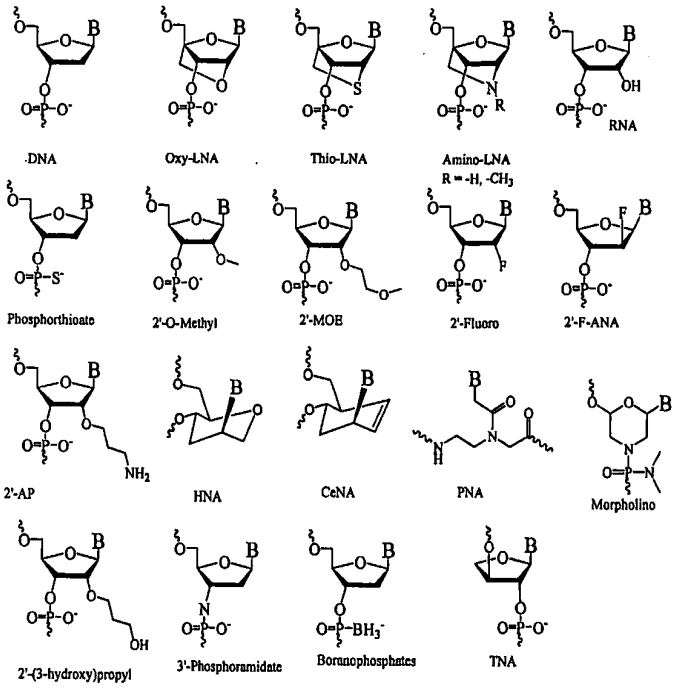
Synthetic Base Pairs



Synthetic purine bases pairing with natural pyrimidines



Suitable examples of backbone units are shown below (B denotes a nucleobase):



The sugar moiety of the backbone is suitably a pentose but may be the appropriate part of an PNA or a six-member ring. Suitable examples of possible pentoses include ribose, 2'-deoxyribose, 2'-O-methyl-ribose, 2'-flour-ribose, and 2'-4'-O-methylene-ribose (LNA). Suitably the nucleobase is attached to the 1' position of the pentose entity.

An internucleoside linker connects the 3' end of preceding monomer to a 5' end of a succeeding monomer when the sugar moiety of the backbone is a pentose, like ribose or 2-deoxyribose. The internucleoside linkage may be the natural occurring phospodiester linkage or a derivative thereof. Examples of such derivatives include phosphorothioate, methylphosphonate, phosphoramidate, phosphotriester, and phosphodithioate. Furthermore, the internucleoside linker can be any of a number of non-phosphorous-containing linkers known in the art.

Preferred nucleic acid monomers include naturally occurring nucleosides forming part of the DNA as well as the RNA family connected through phosphodiester linkages. The members of the DNA family include deoxyadenosine, deoxyguanosine, deoxythymidine, and deoxycytidine. The members of the RNA family include adenosine, guanosine, uridine, cytidine, and inosine.

Selection

Once the library has been formed in accordance with the methods disclosed herein, one must screen the library for chemical compounds having predetermined desirable characteristics. Predetermined desirable characteristics can include binding to a target, catalytically changing the target, chemically reacting with a target in a manner which alters/modifies the target or the functional activity of the target, and covalently attaching to the target as in a suicide inhibitor. In addition to libraries produced as disclosed herein above, libraries prepared in accordance with method A and B below, may be screened according to the present invention.

A. Display molecules can be single compounds in their final "state", which are tagged individually and separately. E.g. single compounds may individually be attached to a unique tag. Each unique tag holds information on that specific compound, such as e.g. structure, molecular mass etc.

B. A display molecule can be a mixture of compounds, which may be considered to be in their final "state". These display molecules are normally tagged individually and separately, i.e. each single compound in a mixture of compounds may be attached to the same tag. Another tag may be used for another mixture of compounds. Each unique tag holds information on that specific mixture, such as e.g. spatial position on a plate.

The target can be any compound of interest. The target can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analog, cofactor, inhibitor, drug, dye, nutrient, growth factor, cell, tissue, etc. without limitation. Particularly preferred targets include, but are not limited to, angiotensin converting enzyme, renin, cyclooxygenase, 5-lipoxygenase, IIL- 1 0 converting enzyme, cytokine receptors, PDGF receptor, type II

Inosine monophosphate dehydrogenase, β-lactamases, and fungal cytochrome P-450. Targets can include, but are not limited to, bradykinin, neutrophil elastase, the HIV proteins, including *tat, rev, gag, int,* RT, nucleocapsid etc., VEGF, bFGF, TGFβ, KGF, PDGF, thrombin, theophylline, caffeine, substance P, IgE, sPLA2, red blood cells, glioblastomas, fibrin clots, PBMCs, hCG, lectins, selectins, cytokines, ICP4, complement proteins, etc.

The upper limit for the strength of the stringency conditions is the disintegration of the complex comprising the displayed molecule and the encoding region. Screening conditions are known to one of ordinary skill in the art.

Complexes having predetermined desirable characteristics can be partitioned away from the rest of the library while still attached to a nucleic acid identifier tag by various methods known to one of ordinary skill in the art. In one embodiment of the invention the desirable products are partitioned away from the entire library without chemical degradation of the attached nucleic acid such that the identifier nucleic acids are amplifiable. The part of the identifier comprising the codons may then be amplified, either still attached to the desirable chemical compound or after separation from the desirable chemical compound.

In a certain embodiment, the desirable display molecule acts on the target without any interaction between the coding sequences attached to the desirable display compound and the target. In one embodiment, the desirable chemical compounds bind to the target followed by a partition of the complex from unbound products by a number of methods. The methods include plastic binding, nitrocellulose filter binding, column chromatography, filtration, affinity chromatography, centrifugation, and other well known methods for immobilizing targets.

Briefly, the library is subjected to the partitioning step, which may include contact between the library and a column onto which the target is bound. All identifier sequences which do not encode for a reaction product having an activity towards the target will pass through the column. Additional undesirable chemical entities (e.g., entities which cross-react with other targets) may be removed by counter-selection methods. Desirable complexes are bound to the column and can be eluted by changing the conditions

of the column (e.g., salt, etc.) or the identifier sequence associated with the desirable chemical compound can be cleaved off and eluted directly.

In a certain embodiment, the basic steps involve mixing the library of complexes with the immobilized target of interest. The target can be attached to a column matrix or microtitre wells with direct immobilization or by means of antibody binding or other high-affinity interactions. In another embodiment, the target and displayed molecules interact without immobilisation of the target. Displayed molecules that bind to the target will be retained on this surface, while nonbinding displayed molecules will be removed during a single or a series of wash steps. The identifiers of complexes bound to the target can then be separated by cleaving the physical connection to the synthetic molecule. It may be considered advantageously to perform a chromatography step after of instead of the washing step. After the cleavage of the physical link between the synthetic molecule and the identifier, the identifier may be recovered from the media and optionally amplified before the decoding step.

In traditional elution protocols, false positives due to suboptimal binding and washing conditions are difficult to circumvent and may require elaborate adjustments of experimental conditions. However, an enrichment of more than 100 to 1000 is rarely obtained. The selection process used in example 7 herein alleviates the problem with false positive being obtained because the non-specific binding complexes to a large extent remain in the reaction chamber. The experiments reported herein suggest that an enrichment of more than 107 can be obtained.

Additionally, chemical compounds which react with a target can be separated from those products that do not react with the target. In one example, a chemical compound which covalently attaches to the target (such as a suicide inhibitor) can be washed under very stringent conditions. The resulting complex can then be treated with proteinase, DNAse or other suitable reagents to cleave a linker and liberate the nucleic acids which are associated with the desirable chemical compound. The liberated nucleic acids can be amplified.

In another example, the predetermined desirable characteristic of the desirable product is the ability of the product to transfer a chemical group (such as acyl transfer) to the target and thereby inactivate the target. One could have a product library where all of

the products have a thioester chemical group, or similar activated chemical group. Upon contact with the target, the desirable products will transfer the chemical group to the target concomitantly changing the desirable product from a thioester to a thiol. Therefore, a partitioning method which would identify products that are now thiols (rather than thioesters) will enable the selection of the desirable products and amplification of the nucleic acid associated therewith.

There are other partitioning and screening processes which are compatible with this invention that are known to one of ordinary skill in the art. In one embodiment, the products can be fractionated by a number of common methods and then each fraction is then assayed for activity. The fractionization methods can include size, pH, hydrophobicity, etc.

Inherent in the present method is the selection of chemical entities on the basis of a desired function; this can be extended to the selection of small molecules with a desired function and specificity. Specificity can be required during the selection process by first extracting identifiers sequences of chemical compounds which are capable of interacting with a non-desired "target" (negative selection, or counter-selection), followed by positive selection with the desired target. As an example, inhibitors of fungal cytochrome P-450 are known to cross-react to some extent with mammalian cytochrome P-450 (resulting in serious side effects). Highly specific inhibitors of the fungal cytochrome could be selected from a library by first removing those products capable of interacting with the mammalian cytochrome, followed by retention of the remaining products which are capable of interacting with the fungal cytochrome.

Enrichment.

The present invention also relates to a method for determining the identity of a chemical entity having a preselected property, comprising the steps of:

i) generating a tagged library of chemical entities by appending unique identifier tags to chemical entities,

ii) subjecting the library to a condition, wherein a chemical entity or a subset of chemical entities having a predetermined property is partitioned from the remainder of the library,

iii) recovering an anti-tag from the partitioned library, said anti-tag being capable of interacting with the unique identifier tag in a specific manner, and

iv) identifying the chemical entity/ies having a preselected function by decoding the anti-tag.

The tag is appended the chemical entity by a suitable process. Notably, each chemical entity is appended a tag by a reaction involving a chemical reaction between a reactive group of the chemical entity and a reactive group of the tag, such as method A and B of the selection section. The attachment of the chemical entity may be directly or through a bridging molecule part. The molecule part may be any suitable chemical structure able to connect the chemical entity to the tag.

The anti-tag has the ability to interact with the unique identifier tag in a specific manner. The chemical structure of the anti-tag is to a large extent dependant on the choice of unique tag. As an example, if the unique tag is chosen as an antibody, the anti-tag is selected as the epitope able to associate with the antibody. In general, it is preferred to use an anti-tag comprising a sequence of nucleotides complementary to a unique identifier tag.

The method may be performed without amplification in certain embodiments. However, when larger libraries are intended, it is in general preferred to use an anti-tag which is amplifiable. Anti-tags comprising a sequence of nucleotides may be amplified using standard techniques like PCR. In the event the anti-tag is a protein, the protein may be amplified by attaching the mRNA which has encoded the synthesis thereof, generating the cDNA from the mRNA and subjecting said mRNA to a translation system. Such system is described in WO 98/31700 the content of which is incorporated herein by reference. An alternative method for amplifying a protein tag is to use phage-displayed proteins.

In the event the tag as well as the anti-tag is a sequence of nucleic acids, a tag:anti-tag hybrid may be formed prior to the subjecting the library to partitioning conditions or subsequent to the partitioning step. In some embodiments of the invention it is preferred to form the tag:anti-tag hybrid prior to the partition step in order to make the appended nucleotide sequence inert relative to the system as it is well known that certain sequences of nucleotides can bind to a target or catalyse a chemical reaction.

The oligonucleotide anti-tag may be formed in a variety of ways. In one embodiment of the invention, the anti-tag is formed as an enzymatic extension reaction. The extension comprises the initial annealing of a primer to the unique identifier tag and subsequent extension of the primer using a polymerase and dNTPs. Other types of extension reactions may also be contemplated. As an example ligases may be used to create the primer starting from di- or trinucleotide substrates and the extension may be performed using a suitable polymerase.

It may be desirable to recover the anti-tag at various steps during the process. To this end it is preferred in some aspects of the invention to provide the primer provided with a handle capable of binding to a suitable affinity partner. An arsenal of different handles and affinity partners are available to the skilled person in the art. The most widely used handle is biotin, which in general are also preferred according to the present invention. Biotin binds to the affinity partner streptavidin or avidin. A standard technique in the laboratory is to recover a biochemical entity having attached a biotin using a solid phase covered with streptavidin. Suitably, the solid phase is a bead which may be separated from the liquid after the binding action by rotation or a magnetic field in case the solid bead comprises magnetic particles.

In other aspects of the present invention, the anti-tag is provided as a separate oligonucleotide. The separate oligonucleotide may be produced using standard amidite synthesis strategies or may be provided using other useful methods. It is in general preferred to provide the oligonucleotide by synthesis, at least in part, because the biotin amidite is easily incorporated in a nascent oligonucleotide strand. Following the addition of an oligonucleotide anti-tag to a liquid comprising chemical entities tagged with complementing oligonucleotide tags a double stranded library is formed as a hybridisation product between the unique identifier tag and the anti-tag oligonucleotide.

As mentioned above, the anti-tag oligonucleotide may be provided with a handle, such as biotin, capable of binding to an affinity partner, such as streptavidin or avidin.

Following the addition of the anti-tag oligonucleotides to the tagged chemical entities, some of the oligonucleotides present in the media may not find a partner. In one aspect of the invention it is preferred that oligonucleotides not hybridised to a cognate unique identifier and/or anti-tag are transformed into a double helix. In other aspects of the

invention single stranded oligonucleotides are degraded prior to step ii) to avoid unintended interference.

The handle may be used to purify the library prior to or subsequent to the partitioning step. In some embodiments of the invention, the purification step is performed prior to the partitioning step to reduce the noise of the system. In another aspect the handle is used to purify the partitioned library subsequent to step ii) in order to recover a double stranded product which may be amplified.

The library is subjected to a condition in order to select chemical entities having a property which is responsive to this condition. The condition may involve the exposure of the library to a target and partitioning the chemical entities having an affinity towards this target. Another condition could be subjecting the library to a substrate and partitioning chemical entities having a catalytical activity relative to this substrate.

The anti-tag can be formed subsequent to the partitioning step. In an aspect of the invention, the single stranded nucleotide serving as a tag is made double stranded while the chemical entity is attached to the target of an affinity partitioning. Optionally, in a repeated temperature cycle, a plurality of anti-tags may be formed as extension products using the tag as template. In another aspect of the invention, the chemical entity bearing the single stranded oligonucleotide is detached from the target and a complementing anti-tag is subsequently prepared.

In the event the anti-tag comprises a handle, this handle can be used to purify the partitioned library. The recovery of the anti-tag is then performed by melting off said anti-tag from a partitioned double stranded library. Optionally, the amount of anti-tags may be multiplied by conventional amplification techniques, such as PCR.

The method according to the invention can be performed using a single partitioning step. Usually, it is preferred, however, to use more than one partitioning step in order to select the candidate having the desired properties from a large library. Thus, the recovered anti-tags may be mixed with the initial library or a subset thereof and the steps of partitioning (step ii)) and recovery (step iii)) may is repeated a desired number of times. Optionally, single stranded moieties in the mixture may be degraded or removed or made inert as described above.

Generally, the partitioned library obtained in step ii) is subjected to one or more further contacting steps using increasing stringency conditions. The stringency conditions may be increased by increasing the temperature, salt concentration, acidity, alkalinity, etc.

In one embodiment of the invention, the partitioned library is not subjected to intermediate process steps prior to a repeated contacting step. Especially, the partitioned library is not subjected to intermediate amplification of the anti-tag. This embodiment may be of advantage when relatively small libraries are used.

The method of the invention terminates with a decoding step, that is a step in which the identity of the chemical entity or entities are deciphered by an analysis of the anti-tag. When the anti-tag is an oligonucleotide, the decoding step iv) may be performed by sequencing an anti-tag nucleotide. Various methods for sequencing are apparent for the skilled person, including the use of cloning and exposure to a microarray.
The tags contain recognizing groups such as e.g. nucleotide sequence(s), epitope(s) a.o. The tags carries information of the entity to which it is attached, such as e.g. entity structure, mass, spatial position (plate information) etc. The tags may be composed of monoclonal antibodies, peptides, proteins, oligonucleotides, DNA, RNA, LNA, PNA, natural peptides, unnatural peptides, polymeric or oligomeric hydrazino aryl and alkyl carboxylic acids, polymeric or oligomeric aminoxy aryl and alkyl carboxylic acids, peptoids, other natural polymers or oligomers, unnatural polymers (molecular weight > 1000 Da) or oligomers (molecular weight < 1000 Da), small non-polymeric molecules (molecular weight < 1000 Da) or large non-polymeric molecules (molecular weight > 1000 Da).

In one preferred embodiment, entities consist of small non-polymeric molecules (molecular weight < 1000 Da). Small molecules are generally the compounds of interest in the quest for drug oral candidates. Especially, small molecules not occurring in Nature are of interest in the drug discovery process and in one aspect of the present invention the method are designed to select a oral drug candidate. A variety of drug candidate libraries are available on the market. The drug candidates of the library usually comprise a reactive group or a group which can be altered into a reactive group. In one preferred aspect of the present invention each of the members of the drug candidate

library is appended a nucleic acid tag via said reactive group of the library member and a reactive group on the nucleic acid. Preferably, the nucleic acid is an oligonucleotide.

In another aspect of the invention, entities consist of large non-polymeric molecules (molecular weight > 1000 Da). In still another embodiment, entities consist of polymeric molecules.

The tags and anti-tags may be composed of RNA linked to monoclonal antibodies, proteins, LNA, PNA, natural polypeptides, unnatural polypeptides, polymeric or oligomeric hydrazino aryl or alkyl carboxylic acids, polymeric or oligomeric aminoxy aryl or alkyl carboxylic acids, other natural polymers or oligomers, unnatural polymers (molecular weight > 1000 Da) or oligomers (molecular weight < 1000 Da), small non-polymeric molecules (molecular weight < 1000 Da) or large non-polymeric molecules (molecular weight > 1000 Da).

Alternatively, anti-tags may be composed of DNA linked to monoclonal antibodies, proteins, LNA, PNA, natural polypeptides, unnatural polypeptides, polymeric or oligomeric hydrazino aryl or alkyl carboxylic acids, polymeric or oligomeric aminoxy aryl or alkyl carboxylic acids, other natural polymers or oligomers, unnatural polymers (molecular weight > 1000 Da) or oligomers (molecular weight < 1000 Da), small non-polymeric molecules (molecular weight < 1000 Da) or large non-polymeric molecules (molecular weight > 1000 Da). Alternatively, anti-tags are just composed of oligonucleotides, DNA or RNA. In a preferred embodiment, anti-tags are composed of DNA. In another preferred embodiment anti-tags are composed of RNA.

Anti-tags which are linked to DNA or RNA are also encoded by the DNA/RNA linked to them, e.g. phage displayed or polysome displayed antibodies, peptides or proteins, and via DNA-templated synthesis of anti-tags, where the DNA encode the synthesis of the anti-tag, which is linked to its DNA during its synthesis.

Each chemical compound or group of compounds may be associated with a tag through formation of a covalent or non-covalent bond. For covalent bond formation, tagging may involve, but is not limited to, the formation of a cycloaddition product, an alkylation product, an arylation product, an acylation product, an amide bond, a carboxylic ester bond, a sulfonamide bond, a disulfide bond, an S-alkyl bond, an NR-alkyl

bond, an O-alkyl bond, an aryl-vinyl bond, an alkyne-vinyl bond, an oxime bond, an imine bond, a bicyclic product, a trizole, a hexene, a 7-Oxa-bicyclo[2.2.1]hept-2-ene derivative, a 7-Aza-bicyclo[2.2.1]hept-2-ene derivative or a 7-Methyl-7-aza-bicyclo[2.2.1]hept-2-ene. Non-covalent bonds may involve, but are not limited to, attachment via e.g. hydrogen bonding, van der Waals interactions, pi-stacking or through hybridization. Hybridization may be between complementary strands of DNA, RNA, PNA or LNA or mixtures thereof. In such case both the tag and the chemical compound carries such a strand complementary to each other. The tagged entity, compound or mixture of compounds may be transformed into a new tagged entity, e.g. by transformation of the entity or by transformation of the tag. The transformation may be caused by either chemical or physical transformations such e.g. addition of reagents (e.g. oxidizing or reducing agents, pH adjustment a.o.) or subjection to UV-irradiation or heat.

The complex between tags and anti-tags may be formed on individually tagged entities immediately after tagging. Alternatively, after mixing individually tagged entities, either before or after the optionally use of library purification, or either before or after library enrichment for specific properties.
When tags and anti-tags are composed of nucleotides the complex consists of a double stranded nucleotide, e.g. duplex DNA or hybrids DNA/RNA.

The purification handle (denoted "@") may be connected to the anti-tag. The purification handle contains a recognizing group(s) such as e.g. nucleotide sequence(s), epitopes, reactive groups, high affine ligands a.o. The purification handles may be composed of monoclonal antibodies, peptides, proteins, DNA, RNA, LNA, PNA, natural peptides, unnatural peptides, polymeric or oligomeric hydrazine aryl or alkyl carboxylic acids, polymeric or oligomeric aminoxy aryl or alkyl carboxylic acids, other natural polymers or oligomers, unnatural polymers (molecular weight > 1000 Da) or oligomers (molecular weight < 1000 Da), small non-polymeric molecules (molecular weight < 1000 Da) or large non-polymeric molecules (molecular weight > 1000 Da). Purification handles may e.g. be a nucleotide sequence, biotin, streptavidin, avidin, "his-tags", mercapto groups or disulfide/activated disulfide groups. The purification handle may be part of the anti-tag, e.g. in the case the anti-tag is nucleotide based or e.g. antibodies where part of the antibody may serve as epitop for another antibody (e.g. immobilized antibody which serve as purification filter).

Purification filters contains components which associate, interact or react with purification handles whereby a complex is formed. This complex allows separation of non-complexed tagged entities and complexed tagged entities. The purification filter contains a recognizing group(s) such as e.g. nucleotide sequence(s), epitopes, reactive groups, high affine ligands a.o. The purification filter may be composed of monoclonal antibodies, peptides, proteins, DNA, RNA, LNA, PNA, natural peptides, unnatural peptides, polymeric or oligomeric hydrazino aryl or alkyl carboxylic acids, polymeric or oligomeric aminoxy aryl or alkyl carboxylic acids, other natural polymers or oligomers, unnatural polymers (molecular weight > 1000 Da) or oligomers (molecular weight < 1000 Da), small non-polymeric molecules (molecular weight < 1000 Da) or large non-polymeric molecules (molecular weight > 1000 Da). Purification filters may e.g. be a nucleotide sequence, biotin, strepdavidin, avidin, "his-tags", mercapto groups or disulfide/activated disulfide groups.

The library is probed and enriched for properties. Properties may be affinity, catalytic activity or membrane penetrating capability a.o.

Amplification may use PCR or RTPCR techniques. Anti-tags are amplifiable in some aspects of the invention. Anti-tags may be separated from tags by use of physical or chemical means, such as e.g. UV-irradiation, heat, pH-adjustment, use of salt solutions a.o.

Isolated tagged entities may be identified either trough their tag or anti-tag. Identification may be accomplished by cloning of anti-tags and sequencing their DNA/RNA or through mass analysis of either tagged entities or anti-tags or complexes of anti-tags/tagged entities.

The library of tagged entities may involve 10-10^{20} or 10-10^{14} or 10-10^{2} or 10-10^{3} or 10^{2}-10^{3} or 10^{2}-10^{4} or 10^{3}-10^{8} or 10^{3}-10^{8} or 10^{3}-10^{10} or 10^{3}-10^{14} or 10^{6}-10^{6} or 10^{5}-10^{8} or 10^{5}-10^{10} or 10^{5}-10^{14} or 10^{8}-10^{14} or 10^{14}-10^{20} entities.

Library complexes of tagged entities and anti-tags may be enriched for properties prior to purification by use of purification handle and purification filter or after purification.

The term unique, when used together with sequences of nucleotides, implies that at least one of the nucleobases and/or backbone entities of the sequence does not appear together with different chemical entities. Preferably, a specific sequence is unique due to fact that no other chemical entities are associated with the same sequence of nucleobases.

Once the library has been formed, one must screen the library for chemical compounds having predetermined desirable characteristics. Predetermined desirable characteristics can include binding to a target, catalytically changing the target, chemically reacting with a target in a manner which alters/modifies the target or the functional activity of the target, and covalently attaching to the target as in a suicide inhibitor.

The target can be any compound of interest. The target can be a protein, peptide, carbohydrate, polysaccharide, glycoprotein, hormone, receptor, antigen, antibody, virus, substrate, metabolite, transition state analog, cofactor, inhibitor, drug, dye, nutrient, growth factor, cell, tissue, etc. without limitation. Particularly preferred targets include, but are not limited to, angiotensin converting enzyme, renin, cyclooxygenase, 5-lipoxygenase, IIL- 1 0 converting enzyme, cytokine receptors, PDGF receptor, type II inosine monophosphate dehydrogenase, β-lactamases, and fungal cytochrome P-450. Targets can include, but are not limited to, bradykinin, neutrophil elastase, the HIV proteins, including *tat, rev, gag, int,* RT, nucleocapsid etc., VEGF, bFGF, TGFβ, KGF, PDGF, thrombin, theophylline, caffeine, substance P, IgE, sPLA2, red blood cells, glioblastomas, fibrin clots, PBMCs, hCG, lectins, selectins, cytokines, ICP4, complement proteins, etc.

The stringency conditions under which the library are screened are normally limited to such condition that maintain the hybridisation between the identifier tag and the anti-tag. High stringency conditions may be applied, however, followed by a renewed synthesis or attachment of the anti-tag. Screening conditions are known to one of ordinary skill in the art.

Chemical compounds having predetermined desirable characteristics can be partitioned away from the rest of the library while still attached to a nucleic acid identifier tag by various methods known to one of ordinary skill in the art. In one embodiment of the invention the desirable products are partitioned away from the entire library without

chemical degradation of the attached nucleic acid such that the identifier nucleic acids are amplifiable. The identifier tag may then be amplified, either still attached to the desirable chemical compound or after separation from the desirable chemical compound.

In the most preferred embodiment, the desirable chemical compound acts on the target without any interaction between the tag attached to the desirable chemical compound and the target. In one embodiment, the desirable chemical compounds bind to the target and the bound tag-desirable chemical compound-target complex can be partitioned from unbound products by a number of methods. The methods include nitrocellulose filter binding, column chromatography, filtration, affinity chromatography, centrifugation, and other well known methods.

Briefly, the library is subjected to the partitioning step, which may include contact between the library and a column onto which the target is bound. All tags which have not formed hybridisation products with a chemical entity-tag aggregate or those tags associated with undesirable chemical entities will pass through the column. Additional undesirable chemical entities (e.g., entities which cross-react with other targets) may be removed by counter-selection methods. Desirable complexes are bound to the column and can be eluted by changing the conditions of the column (e.g., salt, etc.) or the tag associated with the desirable chemical compound can be cleaved off and eluted directly.

Additionally, chemical compounds which react with a target can be separated from those products that do not react with the target. In one example, a chemical compound which covalently attaches to the target (such as a suicide inhibitor) can be washed under very stringent conditions. The resulting complex can then be treated with proteinase, DNAse or other suitable reagents to cleave a linker and liberate the nucleic acids which are associated with the desirable chemical compound. The liberated nucleic acids can be amplified.

In another example, the predetermined desirable characteristic of the desirable product is the ability of the product to transfer a chemical group (such as acyl transfer) to the target and thereby inactivate the target. One could have a product library where all of the products have a thioester chemical group. Upon contact with the target, the desirable products will transfer the chemical group to the target concomitantly changing the

desirable product from a thioester to a thiol. Therefore, a partitioning method which would identify products that are now thiols (rather than thioesters) will enable the selection of the desirable products and amplification of the nucleic acid associated therewith.

There are other partitioning and screening processes which are compatible with this invention that are known to one of ordinary skill in the art. In one embodiment, the products can be fractionated by a number of common methods and then each fraction is then assayed for activity. The fractionization methods can include size, pH, hydrophobicity, etc.

Inherent in the present method is the selection of chemical entities on the basis of a desired function; this can be extended to the selection of small molecules with a desired function and specificity. Specificity can be required during the selection process by first extracting identifier sequences of chemical compounds which are capable of interacting with a non-desired "target" (negative selection, or counter-selection), followed by positive selection with the desired target. As an example, inhibitors of fungal cytochrome P-450 are known to cross-react to some extent with mammalian cytochrome P-450 (resulting in serious side effects). Highly specific inhibitors of the fungal cytochrome could be selected from a library by first removing those products capable of interacting with the mammalian cytochrome, followed by retention of the remaining products which are capable of interacting with the fungal cytochrome.

Following the selection procedure, anti-tags are recovered. The recovery may be performed by subjecting the selected complexes to stringency conditions which will detach the anti-tag sequences from the identifier tag. In the event the tag and the anti-tag are nucleic acids, the stringency conditions may be increased by increasing the temperature gradually until the two strands of the double helix are melted apart. Further copies of anti-tag sequences may be provided by extension of the identifier sequences using a suitable primer and a polymerase. In the alternative, the recovered anti-tag sequence and/or the identifier sequence tag may be subjected to PCR to form a double stranded product. The strands comprising the sequence that complements at least a part of a unique identifier sequence are subsequently isolated.

The selected chemical entity may be attached to the target during the extension or amplification or may be detached from the target. In one aspect of the invention, it is pre-

ferred that the target is immobilised and the chemical compound remain attached to the target during the extension or amplification, to allow for easy recovery of the extension or amplification product by simple elution. In another aspect the selected chemical entities are separated from the unique identifier sequences, prior to, simultaneous with or subsequent to the recovery of the enrichment sequences.

In order to recover the desired anti-tag sequences, it may be appropriate to provide the native as well as the amplified, if present, anti-tag sequences with one part of a molecular affinity pair. The one part of a molecular affinity pair is also referred to herein as a handle. The anti-tags may then be recovered by using the other part of the molecular affinity pair attached to a solid phase, which is possible to isolate. The essential property of the molecular affinity pair is that the two parts are capable of interacting in order to assemble the molecular affinity pair. In the biotechnological field a variety of interacting molecular parts are known which can be used as the molecular affinity pair. Examples include, but are not restricted to protein-protein interactions, protein-polysaccharide interactions, RNA-protein interactions, DNA-DNA interactions, DNA-RNA interactions, RNA-RNA interactions, biotin-streptavidin interactions, enzyme-ligand interactions, antibody-ligand interaction, protein-ligand interaction, ect.

A suitable molecular affinity pair is biotin-streptavidin. The anti-tag sequences can be provided with biotin, e.g. by using a primer attached to a biotin moiety in the amplification or extension step and contacting the biotin tagged anti-tag sequence with beads coated with streptavidin.

After the recovery of the anti-tag sequences, these are contacted with the initial library or a fraction thereof and an enriched library is allowed to be formed by the hybridisation of the anti-tag sequences to the cognate sequence of the unique identifier tag.

The method according to the invention may be repeated one or more times. In a second round of the method, the part of the single stranded library not recognized by an anti-tag sequence may be cleared from the reaction media or the remaining part of the single stranded library may remain in admixture with the enrich library. In general, it is not necessary to separate the remaining part of the single stranded library from the media before the enriched double stranded library is subjected to a second contact with the target because conditions for the preselected function usually are more stringent

than the first round, wherefore the members of the single stranded library presumably will not bind to the target. However, to reduce the noise of the system, it may be useful at some events to withdraw from the media the members of the single stranded initial library not mated with an anti-tag sequence. If the anti-tag sequences are provided with one part of a molecular affinity pair, like biotin, the chemical compounds of interest can be extracted from the media by treatment with immobilized streptavidin, e.g beads coated with streptavidin.

As mentioned above, the conditions for performing the second or further selection step is generally more stringent than in the first or preceding step. The increasing stringency conditions in sequential selection rounds provide for the formation of a sub-library of chemical compounds which is narrowed with respect to the number but enriched with respect to the desired property.

In the present description with claims, the terms nucleic acid, oligonucleotide, oligo, and nucleotides are used frequently. The terms nucleotide, nucleotide monomer, or mononucleotides are used to denote a compound normally composed of two parts, namely a nucleobase moiety, and a backbone. The back bone may in some cases be subdivided into a sugar moiety and an internucleoside linker. Mononucleotides may be linked to each other to form a oligonucleotide. Usually, the mononucleotides are linked through an internucleoside linkage. The term nucleic acid covers mononucleotides as well as oligonucleotides. Usually, however, the term denotes an oligonucleotide having from 2 to 30 mononucleotides linked together through internucleoside linkers.

Determining the coding part of the bifunctional complex

The coding part of the identifier sequence present in the isolated bifunctional molecules or the separated identifier oligonucleotides is determined to identify the chemical entities that participated in the formation of the display molecule. The synthesis method of the display molecule may be established if information on the functional entities as well as the point in time they have been incorporated in the display molecule can be deduced from the identifier oligonucleotide. It may be sufficient to get information on the chemical structure of the various chemical entities that have participated in the display molecule to deduce the full molecule due to structural constraints during the formation. As an example, the use of different kinds of attachment chemistries may ensure that a chemical entity on a building block can only be transferred to a single position on a

scaffold. Another kind of chemical constrains may be present due to steric hindrance on the scaffold molecule or the functional entity to be transferred. In general however, it is preferred that information can be inferred from the identifier sequence that enable the identification of each of the chemical entities that have participated in the formation of the encoded molecule along with the point in time in the synthesis history the chemical entities have been incorporated in the (nascent) display molecule.

Although conventional DNA sequencing methods are readily available and useful for this determination, the amount and quality of isolated bifunctional molecule may require additional manipulations prior to a sequencing reaction.

Where the amount is low, it is preferred to increase the amount of the identifier sequence by polymerase chain reaction (PCR) using PCR primers directed to primer binding sites present in the identifier sequence.

In addition, the quality of the isolated bifunctional molecule may be such that multiple species of bifunctional molecules are co-isolated by virtue of similar capacities for binding to the target. In cases where more than one species of bifunctional molecule are isolated, the different isolated species must be separated prior to sequencing of the identifier oligonucleotide.

Thus in one embodiment, the different identifier sequences of the isolated bifunctional complexes are cloned into separate sequencing vectors prior to determining their sequence by DNA sequencing methods. This is typically accomplished by amplifying all of the different identifier sequences by PCR as described herein, and then using a unique restriction endonuclease sites on the amplified product to directionally clone the amplified fragments into sequencing vectors. The cloning and sequencing of the amplified fragments then is a routine procedure that can be carried out by any of a number of molecular biological methods known in the art.

Alternatively, the bifunctional complex or the PCR amplified identifier sequence can be analysed in a microarray. The array may be designed to analyse the presence of a single codon or multiple codons in an identifier sequence.

Synthesis of nucleic acids

Oligonucleotides can be synthesized by a variety of chemistries as is well known. For synthesis of an oligonucleotide on a substrate in the direction of 3' to 5', a free hydroxy terminus is required that can be conveniently blocked and deblocked as needed. A preferred hydroxy terminus blocking group is a dimexothytrityl ether (DMT). DMT blocked termini are first deblocked, such as by treatment with 3% dichloroacetic acid in dichloromethane (DCM) as is well known for oligonucleotide synthesis, to form a free hydroxy terminus.

Nucleotides in precursor form for addition to a free hydroxy terminus in the direction of 3' to 5' require a phosphoramidate moiety having an aminodiisopropyl side chain at the 3' terminus of a nucleotide. In addition, the free hydroxy of the phosphoramidate is blocked with a cyanoethyl ester (OCNET), and the 5' terminus is blocked with a DMT ether. The addition of a 5' DMT-, 3' OCNET-blocked phosphoramidate nucleotide to a free hydroxyl requires tetrazole in acetonitrile followed by iodine oxidation and capping of unreacted hydroxyls with acetic anhydride, as is well known for oligonucleotide synthesis. The resulting product contains an added nucleotide residue with a DMT blocked 5' terminus, ready for deblocking and addition of a subsequent blocked nucleotide as before.

For synthesis of an oligonucleotide in the direction of 5' to 3', a free hydroxy terminus on the linker is required as before. However, the blocked nucleotide to be added has the blocking chemistries reversed on its 5' and 3' termini to facilitate addition in the opposite orientation. A nucleotide with a free 3' hydroxyl and 5' DMT ether is first blocked at the 3' hydroxy terminus by reaction with TBS-Cl in imidazole to form a TBS ester at the 3' terminus. Then the DMT-blocked 5' terminus is deblocked with DCA in DCM as before to form a free 5' hydroxy terminus. The reagent (N,N-diisopropylamino)(cyanoethyl) phosphonamidic chloride having an aminodiisopropyl group and an OCNET ester is reacted in tetrahydrofuran (THF) with the 5' deblocked nucleotide to form the aminodiisopropyl-, OCNET-blocked phosphonamidate group on the 5' terminus. Thereafter the 3' TBS ester is removed with tetrabutylammonium fluoride (TBAF) in DCM to form a nucleotide with the phosphonamidate-blocked 5' terminus and a free 3' hydroxy terminus. Reaction in base with DMT-Cl adds a DMT ether blocking group to the 3' hydroxy terminus.

The addition of the 3' DMT-, 5' OCNET-blocked phosphonamidated nucleotide to a linker substrate having a free hydroxy terminus then proceeds using the previous tetrazole reaction, as is well known for oligonucleotide polymerization. The resulting product contains an added nucleotide residue with a DMT-blocked 3' terminus, ready for deblocking with DCA in DCM and the addition of a subsequent blocked nucleotide as before.

Brief Description of the Figures

Fig. 1 shows the components of the identifier and the building block
Fig. 2 shows the principle of encoding by extension
Fig. 3 shows the extension region of the building block
Fig. 4 shows the components of the identifier and the building block with internal codons
Fig. 5 shows the principle of encoding by extension with specific annealing
Fig. 6 shows the encoding of scaffolded and polymer molecules
Fig. 7 shows the encoding by extension using three-strand assembly principle
Fig. 8 shows encoding by extension using three-strand assembly principle with specific annealing
Fig. 9 shows the synthesis of three-strand identifier-displayed molecules using a solid-phase approach.
Fig. 10 shows the sequential reaction/extension using platform assembly.
Fig. 11 discloses a general scheme for alternating parallel synthesis of a combinatorial library.
Fig. 12 discloses an encoding method using ligational encoding and a free reactant.
Fig. 13 discloses a library generating method in which a reaction is followed be an encoding step.
Fig. 14 discloses a library generation method using polymerase encoding.
Fig. 15 discloses various embodiments for single encoding methods.
Fig. 16 discloses a double encoding method.
Fig. 17 discloses various double encoding methods.
Fig. 18 discloses encoding using an loop building block.
Fig. 19 discloses a method in which a flexible linker is used in the building block.
Fig. 20 discloses a gel showing the result of an experiment according to example 6.
Fig. 21 discloses a triple encoding method.

Fig. 22 shows the setup used in example 9.
Fig. 23 shows the split-and-mix structure used in example 9.
Fig. 24 discloses an embodiment of library enrichment, amplification and identification.
Fig. 25 shows an embodiment in which anti-tag sequences not hybridised to a identifier sequence are made double stranded and thus inert.
Fig. 26 shows an embodiment in which an enrichment step is before the purification step.
Fig. 27 shows a general principle of library enrichment, amplification, and identification.
Fig. 28 shows a general principle of library enrichment, amplification, and identification omitting the intermediate amplification step between subsequent enrichment procedures.
Fig. 29 shows a general principle of library enrichment, amplification, and identification in which the initial single stranded library is made double stranded prior to enrichment.
Fig. 30 shows a general principle for library enrichment, in which the anti-tag is not formed until after the one and more enrichment processes.
Fig. 31 shows two gels reported in example 13.
Fig. 32 shows the result of the experiment reported in Example 14.
Fig. 33 shows the result of the experiment reported in Example 14.

Detailed description of the figures

Fig. 1 discloses in panel A a hybridisation product between a nascent bifunctional complex and a building block. The nascent bifunctional complex, for short the Identifier, comprises an attachment entity connected to an oligonucleotide identifier region by a linker moiety. The attachment entity may be a single recipient reactive group having been adapted to receive a functional entity or may be a scaffold structure comprising one or more recipient reactive groups. In panel A the attachment entity is indicated as a scaffold having four reactive groups capable of receiving functional entities.

The building block comprises a functional entity attached to an oligonucleotide which is sufficiently complementary to the identifier region to allow for a hybridisation product to be formed. The functional entity is able to be transferred to the attachment entity through a chemical reaction. The complementing identifier region further comprises a unique codon at the 3' or 5' end thereof. The unique codon identifies the functional entity in an unequivocal way.

Following the formation of the hybridisation product between the identifier and the building block, the functional entity and the unique anti-codon are transferred to the identifier. In an aspect of the invention, the linker connecting the functional entity and the complementing identifier region is cleaved simultaneously with the reaction with the attachment entity resulting in a transfer of the functional entity to the attachment entity. Prior to, simultaneously with or subsequent to the transfer, the transcription of the codon occurs. The transcription is performed by an enzyme capable of polymerisation or oligomerisation of oligonucleotides using a template oligonucleotide to form a complementary stand. Usually a polymerase, such as the Pfu polymerase is used together with suitable dNTPs, i.e. a mixture of ATP, CTP, GTP, and TTP, to form the unique codon as an extension of the identifier strand using the unique anti-codon of the building block as template.

Fig. 1, panel B illustrates a typical setup for a second transfer of functional entity. The identifier has been provided with a first functional entity and has been extended by a codon. Furthermore, the codon also comprises a binding region as an extension of the codon. The binding region is usually a constant region transferred to the identifier in the first transfer cycle by the first building block. The identifier forms a hybridisation product with a second building block. The second building block comprises a second functional entity connected to an oligonucleotide sufficient complementary to the identifier region of the identifier to allow for a hybridisation. A part of the complementing identifier region comprises a non-coding region and a region complementing the binding region. The non-coding region opposes the codon transferred in the first cycle and the complementing binding region is complementary to the binding region to allow for a hybridisation which is sufficiently strong for an enzyme to bind to the helix. A second unique anti-codon is attached to the complementary binding region and identifies the second functional entity. The second codon is transferred to the identifier using the second anti-codon as template in the same manner as described above for the first codon.

Fig. 2 illustrates four cycles of functional entity and codon transfer. In the first cycle, a hybridisation product is formed between the identifier and building block. The hybridisation product ensures that the functional entity and the scaffold are brought into close spatial proximity, thus increasing the probability that a reaction will take place. The formation of a duplex between the two oligonucleotides also provides a binding

region for a polymerase. In the presence of a polymerase, a mixture of dNTPs and a suitable puffer such as an aqueous solution containing 20 mM HEPES-KOH, 40 mM KCl and 8 mM $MgCl_2$ and a pH adjusted to 7,4, the unique anti-codon (UA_1) is transferred to the identifier as a codon.

After the transfer of functional entity and codon, respectively, the spent building block is separated from the identifier by increasing the stringency. Usually, the stringency is increased by a increasing the temperature, changing the pH or by increasing the ionic strength. After the rupture of the duple helix structure, the identifier is recovered. In one aspect of the invention the identifier is immobilized to ease the separation from the spent building block. In another aspect the spent building block is degraded chemically or enzymatically. Following the recovery of the identifier a new cycle can be initiated by contacting the identifier with a further building block.

The final product after four cycles of transfer is a bifunctional complex, which comprises a reaction product at one end and an encoding region at the other. The reaction product comprises constituents from the transferred functional entities and the initial scaffold. The encoding region comprises a genetic code for which entities that have been transferred in which order. Thus, the synthetic history may be decoded from the encoding region.

Fig. 3 shows examples of the design of the coding area. Panel A, depicts a detailed view of an example of a design according to fig. 1, panel B. The unique codon transferred in a first cycle is opposed by a partly mis-matching region. To compensate for the decrease in affinity a binding region is following the codon. The binding region is opposed by a matching complementary binding region of the building block.

In Fig. 3, panel B the unique codon incorporated in a first cycle is opposed by a second building block having incorporated in the complementing identifier region a neutral binding region. The neutral binding region is not capable of discriminating between varieties of unique codons, but is able to show some kind of affinity towards the each of the codons. Usually, the neutral binding region comprises one or more universal bases and more preferred the neutral binding region comprises a sequence of universal bases opposing at least a part of the codon region on the identifier.

Fig. 4 shows a hybridisation product between an identifier and a building block wherein the identifier has internal codons and the building block has corresponding anti-codons. The identifier region and the complementing identifier region can also contain specific unique codons and anti-codons, respectively.

The use of internal codons is of particular importance when several rounds of selection are anticipated, especially when the encoded molecule is formed from a PCR product of a previous round. The internal anti-codons in the building block may completely or partly match the identifier sequence or may comprise one or more universal bases to provide for affinity but not for specificity. The role of the internal unique codons is only to guide the annealing between the identifier molecule and the building block molecule. The correct encoding is taken care of by the unique codons which are created in the extension process. These unique codons are passed on to the next generation of molecules and used to decode the synthetic history of the displayed molecules. This system will not be totally dependent on an accurate encoding function by the internal unique codons in order to pass the correct genotype to the next generation of identifier molecules.

In panel A the hybridisation product provides for a spatial proximity between the functional entity and the attachment entity, thus increasing the probability that a reaction occurs. The unique codon templates the codon on the identifier sequence by an enzymatic extension reaction. In panel B a binding region is introduced between each unique coding sequence to provide for affinity of the two strands to each other even though one or more mis-matching bases appear in the codon:non-coding domain of a previously used codon.

Fig. 5 shows an embodiment useful when an amplification step is involved between selections. Initially, a library of complexes is produced as depicted in Fig. 2. The library of the complexes may be subjected to a selection process. The selection process may involve presenting the display molecule on the complex to a target and subsequent selecting the display molecules which shows a desired interaction with the target. It may be advantageously to use relatively mild conditions during the selection process, to obtain a sub-library. The sub-library may be decoded to obtain information on the synthetic history for the entire sub-library. However, it is usually preferred to reduce the sub-library further before a decoding is performed.

The sub-library may be reduced by subjecting it to the target again and use more stringent conditions. However, to obtain a higher number of each of the members of the sub-library before a second selection, it is generally preferred to amplify the complex. Thus, a primer which is loaded with a scaffold is initially annealed to a primer site at one end of the encoding region. Subsequently a transcript is formed. A reverse primer is preferably present to obtain a duple stranded PCR product having a scaffold attached thereto.

This PCR is the basis for the generation of en amplification of the sub-library. The identifier sequence is segregated into a number of internal unique codons, abbreviated IUC in the drawing. The number of the IUCs corresponds to the number of functional entities participating in the formation of the display molecule. The sequence of the IUCs expresses the identity of the individual functional entities and the order of the IUCs indicates the order of reaction of the functional entities. Preferably, a primer region is presented adjacent to the sequence of IUCs to allow for a later amplification of the nucleic acid sequence.

The sub-library is contacted with a plurality of building blocks comprising a transferable functional entity and an internal unique anti-codon (IUA) complementary to at least one of the IUCs. The complementing identifier region is provided with sufficient complementarity to provide for a hybridisation with the oligonucleotide identifier region. In a preferred embodiment the IUCs not identifying a functional entity to be transferred is opposed in the complementary identifier region with a neutral binding region. As mentioned above the neutral binding region may comprise universal bases, i.e. bases that have the ability to be paired with two or more of the naturally occurring nucleobases. Adjacent to the region comprising specific base-pairing sequences and non-specific base-pairing sequences, i.e. the complementary identifier region is a unique anticodon (UA). The UA comprises the same information as the IUA of the complementing identifier region, typically the UA and the IUA has the same sequence on nucleotides.

The transfer step and the reaction step are conducted in several cycles as described above to form a bifunctional complex. In Fig. 5 four cycles are performed, however, it will be appreciated that less than cycles, such as 3 or 2 cycles can be performed to produce a reaction product comprising constituent from 3 or 2 functional entities re-

spectively. Also more, than four cycles may be performed, such as 5 to 20 to form a more diverse library of display molecules. The complexes resulting form the cycles are a reaction product between the functional entities and the scaffold, and an oligonucleotide. The oligonucleotide can be divided into a guiding region, that is, the region that guided the annealing of the individual building blocks, and an encoding region, which comprises the unique codons which have been transferred from the building blocks to the identifier.

Using the above encoding method, allows for the amplification of more and more focused sub-libraries to obtain a sufficient amount of material to allow decoding.

The encoding method shown in Fig. 6 can create both monomer and polymer encoded molecules. Panel A: Complex reaction products can be created using an attachment entity which has reacted with multiple functional entities. Panel B: Polymers can be created using one attachment entity with one reactive group allowing attachment with a functional entity having at least two reactive groups.

Fig. 7 illustrates a three strand assembly procedure for the encoding by extension principle. A: The identifier and building block can be assembled on an assembly platform. This assembly platform contains a unique anticodon region and a unique anticodon where these two elements are directly linked through their sequences. There may be a connecting region linking the unique anticodon region together with the complementing identifier region. B: Describes all the components of the identifier, building block and the assembly platform used in the consecutive reaction, where the identifier also contain a unique codon and a binding region and the assembly platform also contains a non-coding region and a complementing binding region.

In Fig. 8 it is shown that internal codons can also be used for the three-strand assembly principle. This will be useful when selection will be performed in multiple rounds with intermediate amplification steps.

Fig. 9 shows a solid-phase three-strand displayed-molecule synthesis. The assembly platform molecule is attached to a solid support to allow sequential attachment of building blocks to the attachment entity. Different libraries of assembly platform molecules, which is extended with suitable non-coding regions and complementing binding regions, can be used in each step in separate vials. This will allow the use of identical building block and identifier molecules in each step.

Fig. 10 shows the sequential transfer/extension using the assembly platform principle. Each well contains a library of platform molecules. The platform molecule is extended with one unique anticodon in the subsequent wells. A library of identifier and building block molecule is added to the first well which allows specific annealing and transfer of functional entities. The reaction mixture is the transferred to the next wells which finally generates the identifier-displayed library.

Fig. 11 discloses a general scheme for alternating parallel synthesis of combinatorial libraries. In a first step a nascent bifunctional molecule is provided. The nascent bifunctional molecule comprises as one part of the molecule a reactive group, which may appear on a chemical scaffold, and some times referred to herein as a chemical reactive site. Another part of the bifunctional molecule comprises a priming site for addition of a tag. The priming site may be a 3'-OH group or a 5'-phosphate group of a nucleotide in case the tag is a nucleotide. The chemical reactive site and the priming site may optionally be spaced by a linking group. In the event that the linking group is resent it may be a nucleotide or a sequence of nucleotides. The spacing entity may further comprise a hydrophilic linker, such as a polyethylene or polypropylene, to distance the chemical reactive site from the nucleotide. Also comprised in the linking moiety may be a selective cleavable linker that allows the experimenter to separate the display molecule from the coding part.

The nascent bifunctional molecule is divided into a plurality of compartments, usually wells of a microtiter plate or similar equipment that allow easy handling of multiple spatially separated containers. Each of the compartments is reacted with a specific small molecule fragment, also referred to herein as a reactant. Thus, in a first compartment, the nascent bifunctional molecule is reacted with a first small molecule fragment (F_1), in a second compartment; the nascent bifunctional molecule is reacted with a second small molecule fragment (F_2), etc. The number of compartments may in principle be indefinite, however, for practical reasons; the number is usually between 5 and 5000, such as 10 and 500. In each of the compartments the small molecule fragments may be identical or different as the case may be. In each compartment, one, two, or more reactants may participate in the reaction. After the reaction between the

drug fragment and the nascent bifunctional molecule has occurred in each compartment, a tag is added, said tag identifying the small molecule fragment. In certain aspects of the invention, the tag is a nucleic acid. Thus, in the first compartment, a first nucleic acid tag (T_1) is added to the priming site of the reaction product, in the second compartment, a second nucleic acid tag (T_2) is added to the priming site of the second reaction product, etc. Various methods for enzymatic encoding are contemplated and discussed herein. Following the enzymatic addition of the tags in each of the compartments, the contents of the compartments are collected.

In a second round the mixture of bifunctional molecules is split into compartments again. The number of compartments of the second round need not be the same as the number of compartments in the first round. In each compartment the products of the previous round serves as the nascent bifunctional molecule. Thus, a reactive group appearing on the reaction product between the scaffold and the small molecule fragment of the first round is reacted with one or more small molecule fragments of the second round. Thus, in a first compartment, the mixed reaction products of the first round are reacted with a first small molecule fragment (F_1), in a second compartment, the mixed reaction products of the first round are reacted with a second small molecule fragment (F_2), etc. The small molecule fragments F_1, F_2, ... F_x of the second round may be identical or different from the small molecule fragments used in the first round.

After the reactions have been allowed to occur, a tag specifying the small molecule fragment is added. The tag added in the first round usually comprises a priming site that can be used for addition of the tag in the second round so as to produce a linear identifier comprising the tags. In the first compartment, the reacted product is added a first tag which identifies the reactant of the second round that has reacted with the reactive reaction site of the nascent bifunctional molecule; in a second compartment, the product reacted with the second small molecule fragment of the second round is added the tag identifying said reactant, etc. Following the addition of the tags in each compartment, the content of the compartments are mixed in a common pool. The split-reaction-combining cycle can be repeated an appropriate number of times to obtain a library of bifunctional molecules comprising a display molecule part and a coding part. The library may be used in a selection process disclosed elsewhere herein.

Above, the general principle for split-and-mix is disclosed, in which the reaction of the small molecule fragment and the chemical reaction site occurs prior to the encoding step. Obviously, the events can occur in the reverse order or simultaneously.

Fig. 12 schematically shows a 96 well microtiter plate to the left. In each well or in a selected number of wells, the process to the right occurs. Initially, a bifunctional molecule is provided. The bifunctional molecule comprise a chemical reaction site (oval) attached to a codon (rectangle) through a linker (line). To the left of the codon a binding region is provided. Next, a codon oligonucleotide and a splint oligonucleotide are added. The codon oligonucleotide is provided with a codon and flanking binding regions. The splint is designed with sequences complementing the binding region of the nascent bifunctional molecule and a binding region of the codon oligonucleotide such that the ends abut each other under hybridisation conditions. The nascent bifunctional complex, the splint and the codon oligonucleotide forms a hybridisation product under appropriate conditions. A ligase is added to couple the codon oligo to the nascent bifunctional complex. In a second step, a drug fragment, i.e. a reactant, is added and conditions providing for a reaction with the chemical reaction site is instituted.

Then the content of each well is combined and, optionally, divided into a range of wells again for a second round of reaction and encoding. In final step, the combined contents of the wells are used in a selection or partition step, as disclosed herein.

Fig. 13 outlines an embodiment with the encoding and reaction step reversed compared to the embodiment shown in Fig. 12. In a variety of wells a nascent bifunctional complex having a reactive group (Rx) attached to an oligonucleotide (horizontal line) is dispensed. In a first step, the reactive group in each compartment is reacted with a reactant, in a second step a codon oligonucleotide and a splint is added together with a ligase to ligate covalently the codon oligonucleotide to the reacted nascent bifunctional complex, and in a third step the ligation product is recovered. The content of the wells may subsequently be combined and used as a library of bifunctional complexes or recycled for another round of reaction and addition of tag.

Fig. 14 discloses the use of the library produced in accordance Fig. 13, or any other library having a coding part and display molecule part, in a further round. Initially, the

combined contents of the wells from the embodiment of Fig. 13 are dispensed in separate wells. Then an anti-codon oligonucleotide having a binding region which is complementary to the binding region of the nascent bifunctional molecule is added under hybridisation conditions, i.e. conditions which favour the assembly of the hybridisation product between the nascent bifunctional complex and the anti-codon oligonucleotide. Subsequently, or simultaneously with the addition of the anti-codon oligonucleotide, a polymerase, a collection of dNTP (usually, dATP, dGTP, dCTP, and dTTP), and appropriate salts and buffer are added to provide for an extension to occur. The extension (dotted arrow) transcribe the anti-codon to the identifier, thus attaching a tag that encodes the identity of the reactant subsequently reacted at the chemical reaction site. The anti-codon oligonucleotide is connected to a biotin (B) to allow for removal of the oligonucleotide.

Fig. 15 discloses a scheme of various encoding methods combined with a collection of reactants. All the combinations are in according the invention.

Free reactant/polymerase encoding: A nascent bifunctional complex comprises a scaffold (=chemical reaction site) comprising a reactive group and an oligonucleotide part comprising a codon identifying the scaffold. The codon is associated with an oligonucleotide binding region capable of forming a hybridisation product with a complementing binding region of an anti-codon oligonucleotide. The hybridisation product is subjected to an extension reaction, in which the scaffold oligonucleotide is extended over the anti-codon, thereby providing the scaffold oligonucleotide with a codon. Subsequent, simultaneously with or prior to the extension reaction, a free reactant coded for by the anti-codon is reacted with the scaffold.

Zipper Building Block/Polymerase: A nascent bifunctional complex comprises a scaffold (=chemical reaction site) comprising a reactive group and an oligonucleotide part comprising a codon identifying the scaffold. The codon is associated with two oligonucleotide binding region capable of forming a hybridisation product with a complementing binding region of an anti-codon oligonucleotide and a complementing binding region of the reactant. The hybridisation product is subjected to an extension reaction, in which the scaffold oligonucleotide is extended over the anti-codon, thereby providing the scaffold oligonucleotide with a codon. Subsequent, simultaneously with or prior to the extension reaction, a functional entity coded for by the anti-codon is reacted with the scaffold. The selection of polymerase may determine the order of reaction and encoding as some polymerase, such as Sequenase, displaces the binding region

attached to the functional entity, while other polymerases, like Taq polymerase, do not perform the displacement of the binding region. When a zipper building block is used a close proximity between the scaffold and the functional entity is obtained thereby promoting a reaction to take place.

E2 Building Block/Polymerase encoding: A nascent bifunctional complex comprises a chemical scaffold and an oligonucleotide part comprising the codon identifying the scaffold. The oligonucleotide part comprises two binding region on each sides of the codon. An E2 building block anneals to the scaffold oligonucleotide such that the functional entity comes in close proximity as to the scaffold and a double helix is formed just before the anti-codon, thus enable a polymerase to recognize the double helix as a binding area. Applying appropriate conditions and substrates enable the extension of the identifier oligonucleotide over the anti-codon, thus transcribing the genetic information of the function entity to the identifier. Opposing the scaffold codon is a stretch of universal binding nucleotides, such as inosine. Use of an E2 building block allows for one-pot synthesis of a library.

Loop Building block/Polymerase encoding: A nascent bifunctional complex comprises a chemical scaffold and an oligonucleotide part comprising the codon identifying the scaffold. The oligonucleotide part comprises two binding region on each sides of the codon. A loop building block anneals to the scaffold oligonucleotide such that the functional entity comes in close proximity as to the scaffold and a double helix is formed just before the anti-codon, thus enable a polymerase to recognize the double helix as a binding area. Applying appropriate conditions and substrates enable the extension of the identifier oligonucleotide over the anti-codon, thus transcribing the genetic information of the function entity to the identifier. As no sequence on the building block complements the scaffold codon sequence, this codon sequence loops out. Use of a loop building block allows for one-pot synthesis of a library.

N Building Block/Polymerase encoding: A nascent bifunctional complex comprises a chemical scaffold attached to a scaffold codon through a linker. On one or each side of the codon a binding region is present. An N building block comprises a binding region which is complementary to the scaffold binding region and an anti-codon. A functional entity is attached to the codon or a binding region. Under hybridisation conditions the complementary binding regions hybridise and a polymerase extends in both directions, thereby transferring the genetic information of the anti-codon to the oligonucleotide covalently connected to the scaffold. Before, after or simultaneously with the extension reaction, the reaction between the functional entity and the scaffold may take place.

Usually, the functional entity is attached to the anti-codon oligonucleotide *via* a cleavable linker so as to allow for transfer of the functional entity to the scaffold structure.

Free reactant/Ligase: A scaffold entity is attached to an oligonucleotide comprising a codon. The scaffold oligonucleotide further comprises a priming site to which a codon oligonucleotide is ligated. The ligation is performed by a ligase. The ligation can take place in a single stranded or double stranded form. In the single stranded form, a 3' OH (or 5'-phosphate) of the scaffold oligonucleotide is ligated to a 5'-phosphate (or 3'-OH) of the codon oligonucleotide. In the double stranded form, an oligonucleotide complementing the ends of the scaffold and codon oligonucleotides, respectively, is used and designed so that the ends abuts each other. Optionally, the ligation occurs between two double stranded oligonucleotides, i.e. a double stranded scaffold oligonucleotide with an over hang ("sticky end") is ligated to a double stranded codon oligonucleotide provided with a complementing overhang. The type of ligation depends on the selected enzyme. Usually, the double stranded ligation is preferred because the reaction is faster due to the guiding effect of the oligonucleotide complementing the ends. The complementing oligonucleotide is also referred to herein as the splint oligonucleotide. Following, preceding, or simultaneously with the ligation of the codon oligonucleotide to the scaffold oligonucleotide a reaction between the free reactant and the scaffold takes place.

Zipper Building Block/Ligase: A scaffold entity is attached to an oligonucleotide comprising a codon and binding region between the scaffold and the codon. The scaffold oligonucleotide further comprises a priming site to which a codon oligonucleotide is ligated. The ligation is performed by a ligase. The ligation can take place in a single stranded or double stranded form. In the single stranded form, a 3' OH (or 5'-phosphate) of the scaffold oligonucleotide is ligated to a 5'-phosphate (or 3'-OH) of the codon oligonucleotide. In the double stranded form, an oligonucleotide complementing the ends of the scaffold and codon oligonucleotides, respectively, is used and designed so that the ends abuts each other. Optionally, the ligation occurs between two double stranded oligonucleotides, i.e. a double stranded scaffold oligonucleotide with an over hang ("sticky end") is ligated to a double stranded codon oligonucleotide provided with a complementing overhang. The type of ligation depends on the selected enzyme. Usually, the double stranded ligation is preferred because the reaction is faster due to the guiding effect of the oligonucleotide complementing the ends. The complementing oligonucleotide is also referred to herein as the splint

oligonucleotide. A zipper building block is a functional entity attached to a binding oligonucleotide. The binding oligonucleotide is complementing the binding region of the scaffold oligonucleotide, thus forming a hybridisation product under hybridisation conditions. Following, preceding, or simultaneously with the ligation of the codon oligonucleotide to the scaffold oligonucleotide a reaction between the functional entity and the scaffold takes place. The use of the binding region on the reactant ensures a close proximity between the functional entity and the scaffold.

E2 Building Block/Ligational encoding: Initially is provided a nascent bifunctional complex comprising a scaffold attached to an oligonucleotide, said oligonucleotide comprising a codon and a binding region between the scaffold codon and the scaffold codon. The scaffold oligonucleotide also comprises a priming site to which a codon oligonucleotide can be ligated. The scaffold oligonucleotide is hybridised to an E2 building block which carries a double stranded part. The oligonucleotide complementing the anticodon as ligated to the scaffold oligonucleotide using the E2 building block as a template. Before, after or simultaneously with the ligation a reaction takes place between the functional entity and the scaffold.

Loop Building block/Ligational encoding: A bifunctional complex is provided comprising a scaffold attached to an oligonucleotide, wherein the scaffold oligonucleotide comprises a codon flanked by two binding regions. A loop building block is provided which has binding regions complementing the binding regions of the scaffold oligonucleotide. Upon hybridisation, the codon part of the scaffold oligonucleotide loops out. The loop building block also comprises a double stranded codon part. The oligonucleotide complementing the anti-codon part of the loop building block is ligated to the free binding region of the scaffold oligonucleotide. Before, after or simultaneously with the ligation a reaction takes place between the functional entity and the scaffold.

N building block/Ligational encoding: A nascent bifunctional complex is initially provided in which a scaffold via a suitable linker is attached the codon identifying said scaffold or attached to a binding region connect to the codon. A building block having a functional entity connected to a codon is the ligated to the scaffold oligonucleotide to connect the scaffold oligonucleotide with functional entity oligonucleotide. The ligation may be performed in a single stranded or in a double stranded state, depending on the particular enzyme selected for the ligation. Subsequently, the functional entity is reacted with the scaffold. In the alternative, the functional entity and the scaffold are reacted prior to ligation of the respective oligonucleotides.

When a round, i.e. a reaction with and a tagging of the nascent bifunctional complex, has been completed in accordance with any of the above encoding methods, a new round maybe in initialized according to any of the above reaction/encoding methods. Thus, the encoding and reaction in a first round may be the same or different in a subsequent second or further round. A single bifunctional complex or a library of complexes may be generated. When a library is contemplated, one-pot-synthesis can be conducted with the building blocks in which a covalent link between the functional entity and the codon/anti-codon is used, i.e. the columns of E2 building block, loop building block, and N building block. Split and mix synthesis can be performed, when no covalent link between the functional entity/reactant and the codon/anti-codon is present, i.e. in the columns indicating the free reactant and the zipper building block.

Fig. 16 shows a double encoding method, i.e. a method for encoding two or more reactants in one go. In certain embodiments, the multiple encoding methods allow for multi reaction between reactants and scaffold. Initially, a scaffold connected to an oligonucleotide comprising a hybridisation region, a scaffold codon and a binding region is annealed to an E2 building block. Subsequently, an extension is performed in which the anti-codon of the building block is transferred to the identifier. Several polymerases form an overhang of one or more single stranded nucleotides. This overhang is used in the present invention to attach an anti-codon oligo and allow the polymerase to further extent the identifier oligonucleotide over the anti-codon region of the anti-codon oligonucleotide. The transfer of the information of the anti-codon oligonucleotide allows for encoding a third free reactant C. The annealing between the oligonucleotide carrying A and the oligonucleotide carrying B provide for a close proximity between A and B and thus a high local concentration. Thus, when the free reactant C is added a reaction between the three components is favoured. One advantage of double encoding is that it is possible to exchange solvent, such that the reaction not necessarily must take place in the same solvent as the extension occurs.

To the right is illustrated an example, in which the above method is applied on 100 different scaffold oligonucleotides and 100 building blocks. The hybridisation product between the scaffold oligonucleotides and the building block oligonucleotides is divided into 100 different wells. In each of the wells the extension, addition of anti-codon oligonucleotide and reaction with specific free reactant is allowed. In total 10^6 different bifunctional molecules are generated.

Fig. 17 discloses various methods for performing double encoding. In all the examples, the encoding is shown to occur prior to reaction, but it will be within the ambit of the skilled person to perform the reaction first and then the encoding. When a library is contemplated, it is possible to conduct the reaction in a single container (one-pot synthesis) using the N building blocks in combination with any of the encoding methods. For the remaining reactants it is necessary to conduct one or more split-and-mix step. In the combination of the zipper building block, E2 building block, and the loop building block with any of the encoding methods a single split-and-mix step is necessary, whereas two split-and-mix steps are necessary for the free reactant in combination with any encoding method. The scheme makes it possible for the skilled person to select a reaction/encoding method which is useful for a specific reaction. If triple-, quadro-, or multi encoding is contemplated, it is possible to perform such encoding using an embodiment of the double encoding scheme in combination with an embodiment of the single encoding scheme of Fig. 15 one or more times to arrive at an encoding/reaction method that suits the need for a specific chemical reaction.

Fig. 21 discloses a triple encoding method. Initially, a scaffold attached to a scaffold oligonucleotide is provided. The scaffold is attached to a binding region the scaffold oligonucleotide, and the scaffold oligonucleotide is further provided with a codon. The two building blocks of the E2 type is annealed to the scaffold oligonucleotide, thereby bringing the functional entities BB1 and BB2 into close proximity with the scaffold. Simultaneously, prior or subsequent to the addition the building blocks a codon oligonucleotide coding for a third reactant (BB3) is provided which comprises a part complementing a nucleotide sequence of the first building block. The components of the system are allowed to hybridise to each other and a polymerase and a ligase is provided. The polymerase performs an extension where possible and the ligase couples the extended oligonucleotides together so as to form a double stranded product. Following the encoding process, the third reactant is added and conditions are provided which promote a reaction between the scaffold and the reactants. Finally, a selection is used to select reaction products that perform a certain function towards a target. The identifying oligonucleotides of the selected bifunctional complexes are amplified by PCR and identified.

To the right a particular embodiment for carrying out the present invention is indicated. Accordingly, each codon is 5 nucleotides in length and the binding regions flanking the scaffold are 20 nucleotides each. The building blocks designed to hybridise to the binding regions of the scaffold comprises a 20 nucleotide complementing sequence as well as a 5 nucleotide codon.

An embodiment of the enrichment method of the present invention is shown on Fig. 24. Initially, each chemical entity (denoted by letters A, B, C, ..) in a library is attached to a unique identifier tag (denoted a, b, c, ..). The identifier tag comprises information about that particular compound or group of compounds with respect to e.g. structure, mass, composition, spatial position, etc. In a second step, tagged chemical compounds are combined with a set of anti-tag sequences (denoted a', b', c', ..). Each anti-tag sequence carries a handle, like biotin, for purification purposes. The anti-tag sequences comprise a segment which is complementary to a sequence of the identifier sequence. The combination of anti-tag sequences and identifier sequences are allowed to form hybridisation products. Optionally, there may be tagged chemical entities present which have not been recognized by an anti-tag. In a third step, the sequences carrying a handle are removed, i.e. the tagged chemical compounds are left in the media while the matter comprising a handle is transferred to a second media. In the event, the handle is biotin it may be transferred to a second media using immobilized streptavidin.

The purified matter may comprise anti-tag sequences not hybridised to a cognate sequence. As these anti-tag sequences are not coupled to a chemical compound to be selected for, the enrichment sequences may remain in the media. However, in some applications it may be preferably to make the excess anti-tag sequences double stranded, as illustrated in Fig. 25, because the double helix normally is inert relative to the selection procedure. The excess anti-tag sequences may be transformed into the double helix state by the use of a primer together with a suitable polymerase and nucleotide triphosphates.

The purified fraction is in step 4 is subjected to a selection process. The selection comprises probing for a set of properties, e.g. but not limited to affinity for a specific protein. In such a case, entities which do not bind to the specific protein will be eliminated. Anti-tags complexed to entities binding to the specific protein may be recovered/be isolated through e.g. the use of its purification handle.

In step 5 isolated anti-tags are optionally amplified through the use of PCR or RTPCR.

In step 6, the initial library of tagged entities produced in step 1, may undergo further rounds of complexation and screening, i.e. the anti-tags from step 5 may be added the library of tagged entities of step 1 and then be submitted to step 3, step 4 and step 5. Step 6 may be repeated.

In step 7, the isolated anti-tags of step 5 may be cloned and their identity be revealed. E.g. in the case of DNA, sequencing may be applied whereby the identity of specific entities with selected properties in the library of tagged entities will be revealed.

The embodiment shown in Fig. 26 resembles that of Fig. 24 except that the non-complexed components are rendered inert, e.g. if the tags and/or anti-tags are composed of single stranded DNA or RNA, they may be transformed into double stranded DNA, RNA or a hybrid thereof. This may be accomplished by use of a primer, nucleotide triphosphates and a polymerase or transcriptase. Furthermore, the sequence of purification (by use of the purification handle on anti-tags) and probing for properties is changed compared to the method of Fig. 24.

In Fig 27, step 1, a number of entities (denoted by letters A,B,C...), being it mixtures or single compounds are attached to a unique tag more specifically a DNA or RNA sequence or a derivative thereof, holding information on that compound or mixture, such as e.g. structure, mass, composition, spatial information etc.

In step 2, all tags of tagged entities are made double stranded by use of a primer (optionally carrying a @-handle such as e.g. biotin), nucleotide triphosphates and a polymerase or transcriptase. Remaining single stranded DNA or RNA may optionally be digested by use of nucleases.

The mixture, is probed for a set of properties in step 3, e.g. but not limited to affinity for a specific protein. In such a case, entities which do not bind to the specific protein will be eliminated. Anti-tags complexed to entities binding to the specific protein may be recovered/be isolated through e.g. the use of its @-handle.

Isolated anti-tags may optionally be amplified in step 4 through the use of PCR or RTPCR.

In step 5, the library of tagged entities of step 1, may undergo complexation to the isolated and optionally amplified anti-tags of step 3 and 4.

Single stranded components are being digested in step 6 by use of e.g. nucleases. The remaining double stranded subset of the library is optionally subjected to a renewed enrichment of the library according to step 3-6. Steps 3-6 may be repeated as sufficient number of times to obtain an appropriate chemical entity having the desired property.

In step 7, the isolated anti-tags of step 4 can be cloned and their identity be revealed, e.g. in the case of DNA, sequencing may be applied, whereby the identity of specific entities in the library of tagged entities is revealed.

Fig. 28 relates to a method involving a digestion of single stranded oligonucleotides. In a first step a number of entities (denoted by letters A,B,C...), being it mixtures or single compounds, are attached to a unique tag, holding information on that compound or mixture, such as e.g. structure, mass, composition, spatial information etc.

In step 2, mixtures of tagged entities are combined with a set of complementary anti-tags. Anti-tags may be, but is not limited to nucleotide derivatives. Anti-tags may optionally carry a @-handle. The tag and the anti-tags are allowed to form a complex. The complexation may be, but is not limited to hybridization. Some anti-tags will not form a complex with a tagged entity and some tagged entities will not form a complex with an anti-tag.

Non-complexed components is digested in step 3 using e.g. nucleases when the tags and/or anti-tags are composed of DNA or RNA or hybrids thereof.

The mixture of step 3, is probed for a set of properties in step 4, e.g. but not limited to affinity for a specific protein. In such a case, entities which do not bind to the specific protein will be eliminated. Anti-tags complexed to entities binding to the specific protein may be recovered/be isolated through e.g. the use of its @handle. Step 4 may be repeated one or more times.

Isolated anti-tags may optionally be amplified through the use of PCR or RTPCR as illustrated in step 5. Anti-tags may then also be used as described in Figures 24-27.

The isolated anti-tags may be cloned and their identity be revealed in step 6, e.g. in the case of DNA, sequencing may be applied, whereby the identity of specific entities in the library of tagged entities will be revealed.

According to Fig. 29, step 1, a number of entities (denoted by letters A,B,C...), being it mixtures or single compounds, are attached to a unique tag more specifically a DNA or RNA sequence or a derivative thereof, holding information on that compound or mixture, such as e.g. structure, mass, composition, spatial information etc.

All tags of tagged entities are made double stranded in step 2 by use of a primer (optionally carrying a @-handle such as e.g. biotin), nucleotide triphosphates and a polymerase or transcriptase. Remaining single stranded DNA or RNA may optionally be digested by use of e.g. nucleases.

In step 3, the mixture is probed for a set of properties, e.g. but not limited to affinity for a specific protein. In such a case, entities which do not bind to the specific protein will be eliminated. Anti-tags complexed to tags having appended entities binding to the specific protein may be recovered/be isolated through e.g. the use of its @-handle. Step 3 may be repeated one or more times.

According to step 4, isolated anti-tags may optionally be amplified through the use of PCR or RTPCR. Anti-tags may then also be used as described in Figs. 24-27.

The isolated anti-tags may be cloned in step 5 and their identity be revealed, e.g. in the case of DNA, sequencing may be applied. Whereby, the identity of specific entities in the library of tagged entities will be revealed.

Fig. 30, step 1, produces a number of entities (denoted by letters A,B,C...), being it mixtures or single compounds which are attached to a unique tag more specifically a DNA or RNA sequence or a derivative thereof, holding information on that compound or mixture, such as e.g. structure, mass, composition, spatial information etc.

In step 2, the mixture is probed for a set of properties, e.g. but not limited to affinity for a specific protein. In such a case, entities which do not bind to the specific protein will be eliminated. Step 2 may be repeated.

All tags of tagged entities are made double stranded in step 3 by use of a primer (optionally carrying a @-handle such as e.g. biotin), nucleotide triphosphates and a polymerase or transcriptase. Remaining single stranded DNA or RNA may optionally be digested by use of e.g. nucleases.

Anti-tags complexed to tags of entitles binding to the specific protein may be recovered/be isolated in step 4 through e.g. the use of its @-handle. Anti-tags may optionally be amplified through the use of PCR or RTPCR. Anti-tags may then also be used as described in Figs. 24-27.

The isolated anti-tags may be cloned in step 5 and their identity be revealed, e.g. in the case of DNA, sequencing may be applied, whereby, the identity of specific entities in the library of tagged entitles is revealed.

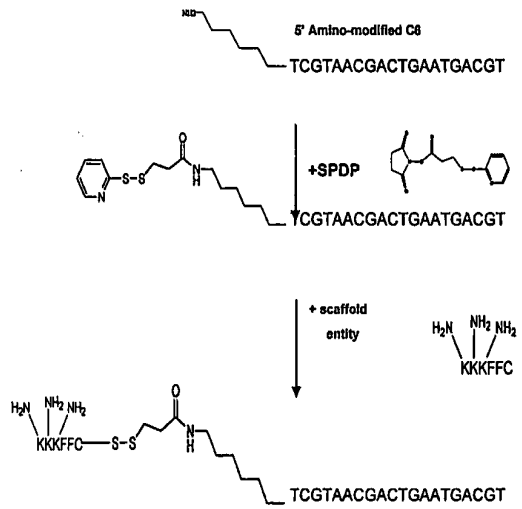
EXAMPLES

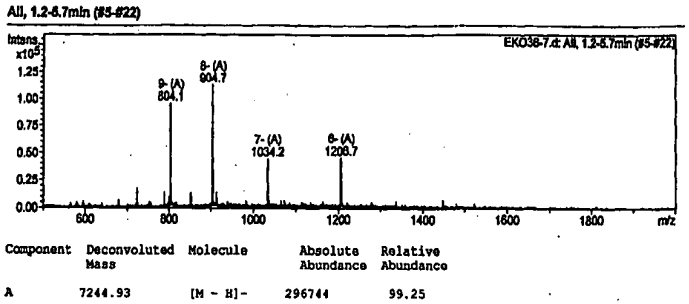
Example 1: Loading of a scaffold onto identifier molecules

An amino-modifier C6 5'-labeled identifier oligo (5'-X-TCGTAACGACTGAATGACGT-3', wherein X may be obtained from Glen research, cat. # 10-1039-90) was loaded with a peptide scaffold (Cys-Phe-Phe-Lys-Lys-Lys, CFFKKK) using SPDP activation (see below). The SPDP-activation of amino-oligo was performed using 160 µl of 10 nmol oligo in 100 mM Hepes-KOH, pH=7.5, and 40 µl 20 mM SPDP and incubation for 2 h at 30°C. The activated amino-oligo was extracted 3 times with 500 µl EtOAc, dried for 10 min in a speed-vac and purified using micro bio-spin column equilibrated with 100 mM Hepes-KOH. The loading of scaffold was then performed by adding 10 µl of 100 mM attachment entity and incubating overnight at 30°C.

The loaded identifier oligo was precipitated with 2 M NH_4OAc and 2 volume 96% ethanol for 15 min at 80°C and then centrifuged for 15 min at 4°C and 15.000g. The pellet

was re-suspended in water and the precipitation was repeated. Wash of the oligo-pellet was done by adding 100 µl of 70% ethanol and then briefly centrifuged. The oligo was re-dissolved in 50 µl H_2O and analysed by MS. The MS analysis was performed after 100 pmol oligo in 10 µl water was treated with 10 µl of ion exchanger resin and incubated minimum 2 h at 25 °C on a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed using a mass spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass, as can be seen below, was 7244.93 Da, which correspond well with the calculated mass, 7244.00 Da. This experimental data exemplify the possibility to load scaffolds onto identifier oligonucleotides. This loaded identifier molecule can be used to receive functional entities from building blocks. This particular scaffold harbours three identical reactive groups, i.e. the amine group of the lycin side chain, and can therefore be transferred with one, two, or three functional entities, which is capable of reacting with the amine groups.



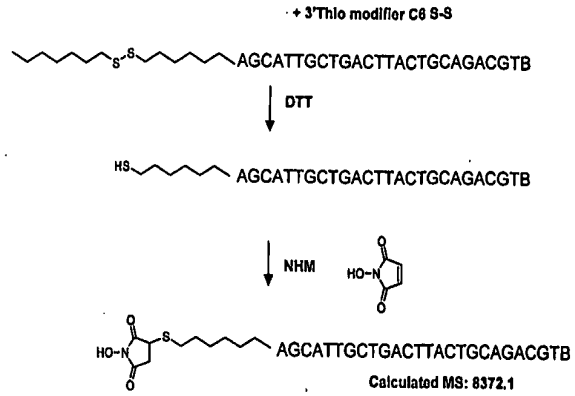


Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	7244.93	[M - H]-	296744	99.25

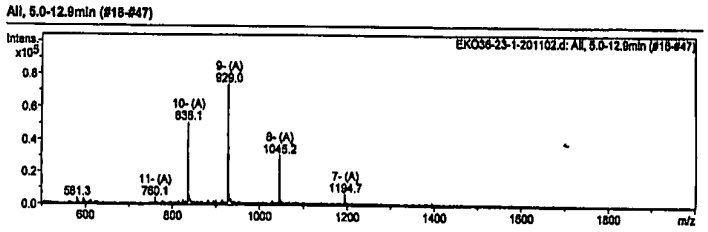
Example 2: Loading of functional entities onto building blocks

Loading of functional entities onto building block molecules can be done using a thiol-oligo (see below). An Biotin 5' labeled and thio-modifier C6 S-S (obtainable from Glen Research, cat # 10-1936-90) 3'-labeled building block oligo (5'-BTGCAGACGTCATTCAGTCGTTACGA-3') was converted to an NHS-oligo using NHM.

10 nmol oligo was dried in speed-vac, re-dissolved in 50 µl 100 mM DTT, 100 mM sodium-phosphate pH 8.0 and incubated at 37 °C for 1 hour. The thiol-oligo was then purified using micro bio-spin column equilibrated with 100 mM Hepes-KOH, pH 7.5. The thiol-oligo was converted to NHS-oligo by adding 100 mM NHM in 100 mM Hepes-KOH pH. 7.5. The sample was incubated at 25 °C over night. The NHS-oligo was then purified using bio-spin column equilibrated with MS-grade H_2O.

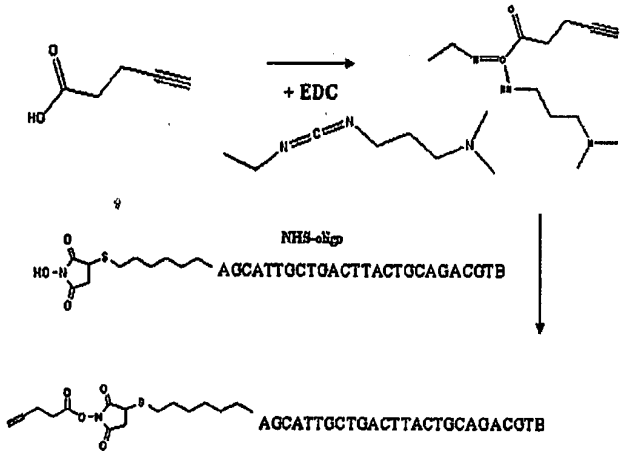


The MS analysis was performed after 100 pmol oligo in 10 µl water was treated with 10 µl of ion exchanger resin and incubated minimum 2 h at 25 °C on a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed using a mass spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass as can be seen below was 8369.32, which correspond well with the calculated mass, 8372.1. The experimental data exemplify the possibility to convert the attachment entity on building block oligonucleotides. This product can later be used to attach transferable functional entities.



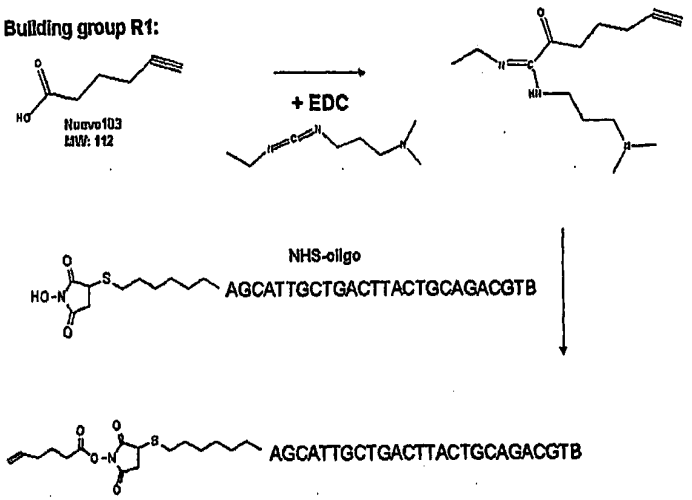
Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	8369.32	[M - H]-	160905	96.87

The NHS-oligo was then used to load functional entities. EDC activation of the functional entity (4-pentynoic acid) was performed mixing 50 µl of 200 mM functional entity in DMF with 50 µl of 200 mM EDC in DMF and incubated for 30 min at 25 °C on a shaker. The loading was then performed using 1 nmol NHS-oligo lyophilized in a speed-vac and 10 µl of the activated building block (see below). This was incubated at 25 °C for 5 min and then mixed with 30 µl 100 mM MES pH. 6.0. The loaded NHS-oligo was purified using bio-spin column equilibrated with 100 mM MES pH 6.0. The loaded building block oligo is then used immediately for the transfer reaction without any MS analysis. This is due to the unstable structure of the functional entity during the conditions used for the MS measurements.



This experiment exemplifies a complete loading of a functional entity onto a building block molecule ready for transfer to an recipient reactive group when annealed to the complementary identifier molecule.

Another example of a functional entity that can be loaded as described above onto a building block is a 5-hexynoic acid as shown below. Again, no MS analysis was performed on this compound due to the unstable structure of the functional entity in the conditions used in the MS measurements.



Example 3: Transfer of functional entities from the building block to the identifier molecule

The attachment entitiy (AE) in the following experiments are either a scaffold, e.g. the peptide, CFFKKK, loaded on an Identifier as prepared in Example 1 or a recipient reactive group exemplified by an amino modified oligonucleotide used as starting material in Example 1. These attachment entities allow transfer of three or one functional entities, respectively.

The identifier used in this experiment is an identifier oligonucleotide loaded with CFFKKK as described in Example 1. The functional entity (FE) in this experiment is the 4-Pentynoic acid, the loading of which was described in Example 2. The identifier molecule loaded with the scaffold is annealed to the loaded building block molecule to bring the attachment entity and the functional entity in close proximity. The annealing is directed by the identifier region in the identifier molecule and the complementary sequence in the building block molecule.

AE-TCGTAACGACTGAATGACGT

\+

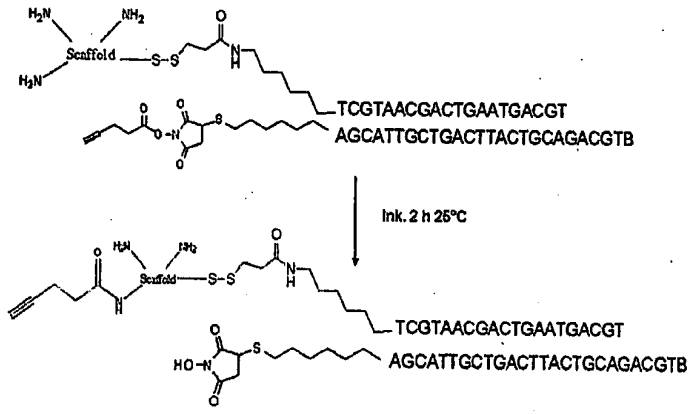
FE-AGCATTGCTGACTTACTGCAGACGTB

AE-TCGTAACGACTGAATGACGT

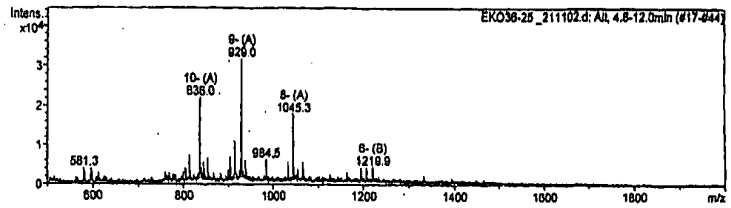
FE-AGCATTGCTGACTTACTGCAGACGTB

After the annealing step between the identifier and building block molecules, the transfer reaction takes place where the functional entity is transferred to the identifier molecule.

The annealing was performed using 600 pmol of the building block and 400 pmol identifier molecules in 0.1 M MES buffer at 25°C in a shaker for 2 hours. The reactive part (functional entity) of the building block was transferred to the one of the amino group on the attachment entity on the identifier molecule during the annealing (see below). After annealing the sample was purified by micro-spin gel filtration and analyzed by MS. The sample was prepared for MS analysis using equal amount of sample (about 100 pmol) and ion exchanger resin and incubated minimum 2 h at 25° in a shaker. After incubation the resin was centrifuged down and 15 µl of the supernatant was added 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 ul acetonitrile. The sample was analysed on a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass (see below) was 7323.45 Da, which correspond well with the calculated mass, 7324.00 Da. Thus, the MS spectrum of the identifier molecule after the transfer reaction shows a mass corresponding to the transferred functional entity on the identifier molecule.



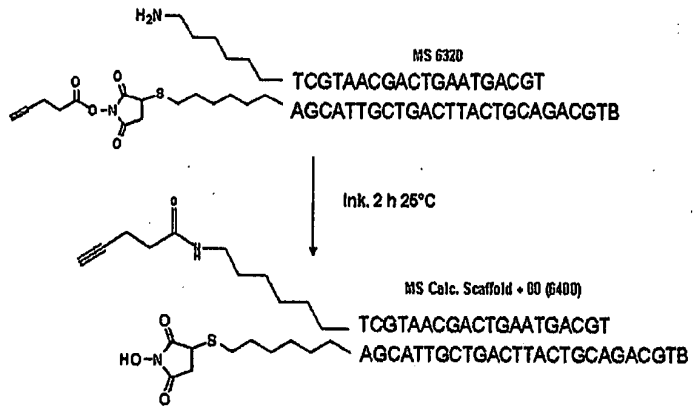
All, 4.8-12.0min (#17-#44)



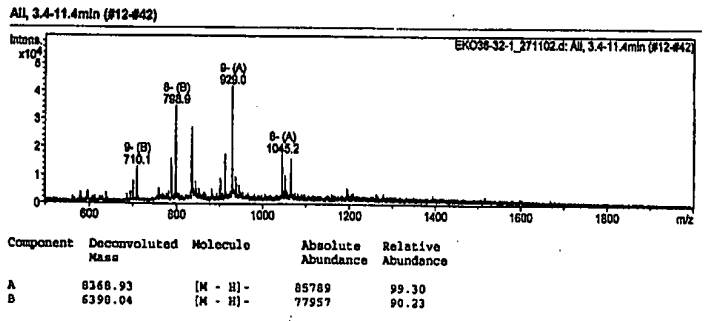
Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	8369.18	[M - H]-	75183	99.24 → NH5-Oligo
B	7323.45	[M - H]-	38073	50.26 → TBAMGB
C	7244.34	[M - H]-	20775	27.42 → SCAFFOLD

Another example of transfer of functional entity is shown below using the amino oligo directly as the AE on the identifier molecule. The functional entity on the building block molecule used in this experiment was 4-pentynoic acid, as disclosed in example 2.

The annealing was performed using 500 pmol of the building block and the identifier molecules in 0.1 M MES buffer and incubating the mixture at 25°C in a shaker for 2 hours. The reactive part (functional entity) of the building block was transfer to the amino group on the identifier molecule during the annealing (see below). After annealing and transfer the sample was purified by micro-spin gel filtration and analyzed by MS. The sample was prepared for MS analysis using equal amount of sample (about 100 pmol) and ion exchanger resin and incubated minimum 2 h at 25° in a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was added 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 ul acetonitrile.



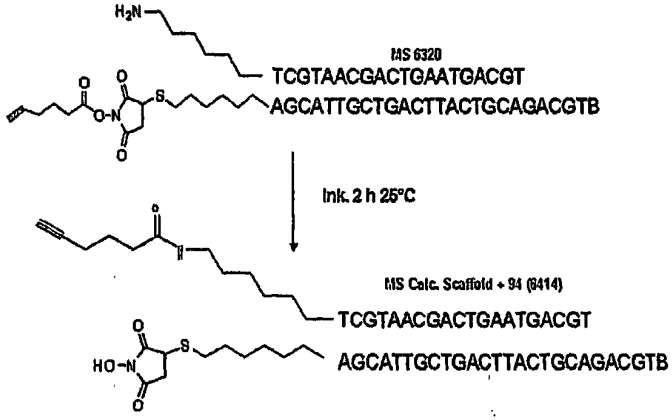
The sample was analysed on a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass was 6398.04 Da, which correspond well with the calculated mass, 6400.00 Da. Thus, the MS spectra of the identifier molecule after transfer of the functional entity show a mass corresponding to the transferred functional entity on the identifier molecule. This example shows that functional entities can be transferred using this setup of a building block molecule and an identifier molecule.



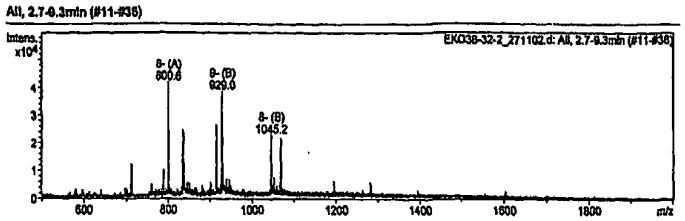
Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	8368.93	[M - H]-	85789	99.30
B	6398.04	[M - H]-	77957	90.23

Another example of transfer of functional entity is shown below using the amino oligo directly as the identifier molecule. The functional entity used in this experiment was 5-Hexynoic acid, prepared as shown in example 2.

The annealing was performed using 500 pmol of the building block and 500 pmol of the identifier molecules in 0.1 M MES buffer incubated at 25°C in a shaker for 2 hours. The reactive part (functional entity) of the building block was transferred to the amino group on the identifier molecule (see below). After annealing and transfer the sample was purified by micro-spin gel filtration and analyzed by MS. The sample was prepared for MS analysis using equal amount of sample (about 100 pmol) and ion exchanger resin and incubated minimum 2 h at 25 °C in a shaker. After incubation the resin was removed by centrifugion and 15 µl of the supernatant was added 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 ul acetonitrile.



The sample was analysed on a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus). The observed mass was 6411.96 Da, which correspond well with the calculated mass, 6414 Da. Thus, the MS spectra of the identifier molecule after transfer of the functional entity show a mass corresponding to the transferred functional entity onto the identifier molecule. This example shows that functional entities can be transferred using this setup of a building block molecule and an identifier molecule.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	6411.96	[M - H]-	86999	98.85
B	8368.87	[M - H]-	84433	93.78

Example 4: Extension of the identifier molecule to transfer unique codons

After the transfer of the functional entity (FE) to the attachment entity (AE) on the identifier molecule, the identifier molecule is extended in order to transfer the unique codon, that identifies the transferred functional entity, to the identifier molecule. This is accomplished by adding a suitable polymerase and a polymerase buffer containing the wild type nucleotides (dATP, dTTP, dCTP, dGTP). This will extend the identifier molecule in the 3'-end towards the end of the 5'-end of the building block molecule.

The extension of the identifier molecule to transfer the unique anticodon(s) is preferably performed after the transfer of the FE as shown below.

FE-AE-TCGTAACGACTGAATGACGT

-AGCATTGCTGACTTACTGCAGACGTB

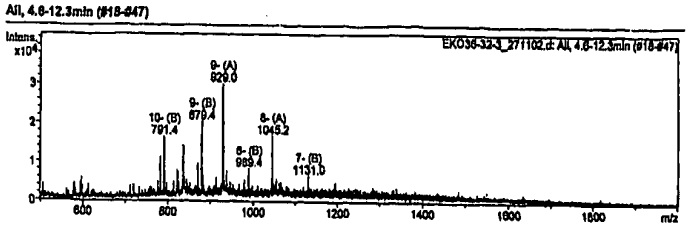
↓

FE-AE-TCGTAACGACTGAATGACGT**CTGCT**

-AGCATTGCTGACTTACTGCAGACGTB

The extension was performed using 15 units Taq polymerase in a buffer containing 0.4 mM of each nucleotide in an extension buffer (20 mM HEPES-KOH, 40 mM KCl, 8 mM $MgCl_2$, pH=7,4). After the extension reaction the sample was analyzed using MS. The MS analysis was performed using about 100 pmol purified extension mixture in a half volume of ion exchanger resin and incubated minimum 2 h at 25 °C in a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed on a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus).

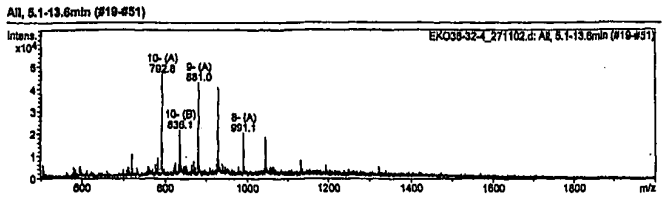
The MS data for extension on the identifier molecule with a transferred 4-Pentynoic acid is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	8368.79	[M - H]-	60417	98.68
B	7922.53	[M - H]-	51334	83.85

The observed mass was 7922.53 Da, which correspond well with the calculated mass, 7924.00 Da. The MS spectra of the identifier molecule after the transfer reaction of the functional entity and extension reaction of the encoding region (the unique codon) showed a mass corresponding to the transferred functional entity and the extension on the identifier molecule. This example shows that functional entities can be transferred using this setup with a longer building block molecule than the identifier molecule and that the identifier molecule can be extended using a polymerase after the transfer process. This shows the possibility to transfer both the functional entity and the unique codon from the same building block to an identifier molecule.

Another example showing transfer and extension is for the building block with the functional entity 5-Hexynoic acid. The MS data for extension on the identifier molecule with a transferred 5-Hexynoic acid is shown below.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	7936.99	[M - H]-	107170	97.86
B	8369.04	[M - H]-	79840	72.92

The observed mass was 7936.99 Da, which correspond well with the calculated mass, 7938.00 Da. The MS spectra of the identifier molecule after transfer reaction of the functional entity and extension reaction of the encoding region (the unique codon) showed a mass corresponding to the transferred functional entity and the extension on the identifier molecule. This example also shows that functional entities can be transferred using this setup with a longer building block molecule than the identifier molecule and the identifier molecule can be extended using a polymerase after the transfer process. This exemplifies the possibility to transfer both the functional entity and the unique codon from one building block molecule to one identifier molecule.

Example 5: Library design

The identifier molecule can be designed to operate optimal under various conditions. However, it should contain a few elements that are vital for the function. The identifier molecule should comprise of a sequence that can anneal to the building block and an attachment entity that can accommodate various functional entities. Below is an example on how an identifier molecule can be designed in the extension region. The region that becomes extended during each step of transfer and encoding can be designed using various approaches. Importantly, there must be a base-pair match between the building block and the identifier to allow efficient extension using a polymerase. This can be accomplished using either a region that is constant, the binding region as described in Figure 3 (A), or a region that allow binding to any given sequence, also shown in Figure 3 (B). A combination of these to approaches can also be used.

The first step in the extension process needs no special binding region due to the match of the identifier and the building block molecules (step 1 shown below). However, the subsequently steps needs a binding region sufficient complementary to the identifier molecule to allow for hybridisation because the enzyme, preferably a polymerase must be able to bind to the douplex and perform an extension. The example below shows four steps in the encoding procedure. This process of extension can be continued to obtain the suitable number of transfer of building blocks. The binding region in this example contains 6 nucleotides, but this can be varied dependent on the design of the building blocks.

A possibility to accommodate the possible mismatches in the previous anticodon is to use universal nucleobases, i.e. a nucleobases with the ability to base pair with more than one of the natural nucleobases. A possible base is inosine which can form base pairs with cytidine, thymidine, and adenosine (although the inosine:adenosine pairing presumably does not fit quite correctly in double stranded DNA, so there may be an energetic penalty to pay when the helix bulges out at this purine:purine pairing). In principle, any design that allows extension of the unique codons is possible to use.

The identifier and building blocks:

Identifier:

GCA CAC ATG CAT GAG CAC AC G

Building block library to step 1:

CGT GTG TAC GTA CTC GTG TG CGT GTG NNNNNN TGA CTA

Building block library to step 2:

CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA NNNNNN TGC AAC

Building block library to step 3:

CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC NNNNNN ACT TTG

Building block library to step 4:

CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC IIIIII ACT TTG NNNNNN GAA TTC GGC AAT ACG CAT TAC CG

EcoRI

N: A nucleobase selected from A, G, T, C.

I: Inosine

SUBSTITUTE SHEET (RULE 26)

Example of encoding and extension of the encoding region by transfer of the unique codons in each step:

1.

GCA CAC ATG CAT GAG CAC AC G
CGT GTG TAC GTA CTC GTG TG CGT GTG TCGATG TGA CTA

2.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT
CGT GTG TAC GTA CTC GTG TG CGT GTG TCGATG TGA CTA

3.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA CAATCG TGC AAC

4.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA CAATCG TGC AAC

5.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC CTCTGT ACT TTG

6.

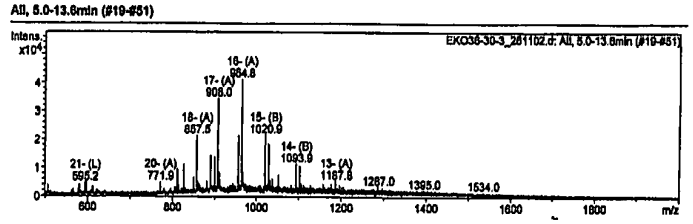
SUBSTITUTE SHEET (RULE 26)

```
GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG GAGACA TGA AAC
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC CTCTGT ACT TTG

7.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG GAGACA TGA AAC
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC IIIIII ACT TTG TAAGCT [redacted] GGC AAT ACG CAT TAC CG

                    EcoRI

8.

GCA CAC ATG CAT GAG CAC AC GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG GAGACA TGA AAC ATTCGA [redacted] CCG TTA TGC GTA ATG GC
CGT GTG TAC GTA CTC GTG TG CGT GTG IIIIII TGA CTA IIIIII TGC AAC IIIIII ACT TTG TAAGCT [redacted] GGC AAT ACG CAT TAC CG
```


The extension was performed using 60 pmol of primer and template in an extension buffer (20 mM HEPES-KOH, 40 mM KCl, 8 mM $MgCl_2$, pH=7,4) and 10mM DTT. For extension in this experiment, 20 U AMV-RT (Promega M510B) was used.

The MS analysis was performed using about 100 pmol extension reactions in half volume of ion exchanger resin and incubated minimum 2 h at 25°C on a shaker. After incubation the resin was removed by centrifugation and 15 µl of the supernatant was mixed with 7 µl of water, 2 µl of piperidine and imidazole (each 625 mM) and 24 µl acetonitrile. The sample was analysed using a Mass Spectroscopy instrument (Bruker Daltonics, Esquire 3000plus).

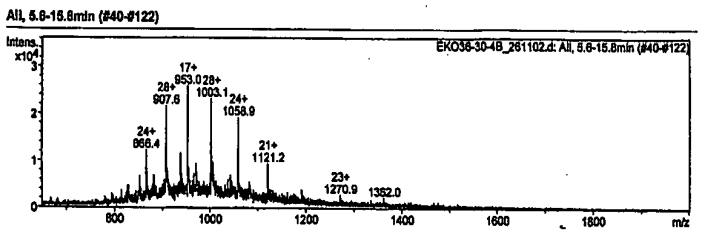
Extension of the primer/template combination in step 3 as show in the example above is shown in data MS graph below. The analysis shows both the mass of the template, 15452.14 Da (expected 15453.86 Da) and the extended primer, 15328.92 (expected 15331.90 Da). No mass for the non-extended (or partially extended) primer was identified indicating a complete extension of the primer. This experiment exemplifies the possibility to transfer the codon to an identifier molecule even when the anticodon is preceded by another anticodon.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance	
A	15452.14	[M - H]-	161269	100.00	- Exo46
B	15328.92	[M - H]-	89333	55.39	- Ext.

In a separate experiment, extension of the primer/template combination in step 7 as described above was examined. The MS data is shown in the graph below. The data shows the mass for the extended primer, 28058.14 Da (expected 28052.30 Da). Again, no mass for the non-extended (or partially extended) primer was identified indicating a complete extension of the primer. This experiment exemplifies the possibility to transfer the codon to an identifier molecule even when the anticodon is

preceded by multiple anticodons. Thus, a complete process of making the encoding region containing the unique codons is feasible for a library.



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	16184.59	[M + H]+	48389	66.47
B	28058.14	[M + H]+	64120	88.08

In conclusion, these experiments show that the polymerase can extend the unique anticodon sequence when using an adjacent unique codon with a helix comprising inosines. This will allow the transfer of the unique anticodons to the identifier molecule in each step of transfer of the functional entities. This experiment shows the possibility to use a binding region after the anticodon region preceding the anticodon that is to be extended in the encoding process. The same approach can be used in the consecutive steps to allow the encoding of a molecule with multiple functional entities attached to the attachment entity.

After the library has been generated and the first selection round has been performed, the selected identifier molecules can be used as a source for the next round of libraries. The selected identifier molecules can be used in subsequently rounds of selection using for example PCR amplification and restriction enzyme digestion as shown below.

PCR product of a new identifier molecule:

GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG GAGACA TGA AAC ATTCGA CAA TTC CCG TTA TGC GTA ATG GC

Cut with EcoRI to obtain the new identifier molecule:

GCA CAC AGCTAC ACT GAT GTTAGC ACG TTG GAGACA TGA AAC ATTCGA C

This new identifier molecule will contain unique codons that encodes for selected displayed molecules in the previous round of selection. This identifier molecule will guide the assembly of the next library to obtain a library that has preferred functional entities. However, the correct encoding will still be determined by the extension process.

Example 6. Flexible linker and loop-out structure in the encoding by extension procedure.

The encoding process can be designed to allow the formation of a loop-out region in the identifier molecule. The encoding process can also be performed using a flexible linker between the complementary identifier region and the complementary binding region.

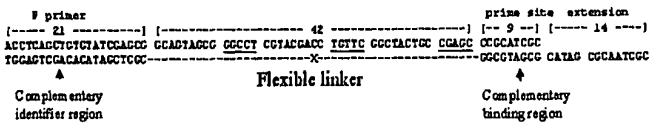
The loop-out strategy is shown in Fig. 18 where the identifier region anneals to the complementary identifier region and the binding region anneals to the complementary binding region. This will form a stretch of single-stranded nucleotides that is not directly participating in the annealing process. This annealing will allow for the transfer by extension of the anticodon region and preferably another binding region that can be used in the next round of extension. The binding region should be long enough to ensure correct annealing and productive extension. The extension will be incomplete if the binding region is to short. It is within the capabilities of the skilled person by simple trial and error experiments to determine the length of the binding region. Usually 5 to 7 nucleotides are sufficient for the binding region.

Another example is to use a flexible linker between the complementary identifier region and the complementary binding region. This is shown in Fig. 19. The identi-

fier region will ensure efficient annealing of the building block to the identifier. The flexible linker will then make sure that the complementary binding region anneals to the binding region to allow extension. The linker can be any type of chemical structure that allow space between the complementary binding region and the complementary binding region, for example, polyethylene glycol (PEG), polyamines, polynucleotides (*e.g.* DNA,RNA, LNA) and polycarbohydrates. The linker length can be varied but a simultaneous annealing of the identifier region and the binding region must be possible.

The setup using a flexible linker was tested using different PEG linkers and different length of the complementary binding region. The PEG linkers (space phosphoramidite 9 and 18) used in this example was obtained from Glen Research (cataloque # 10-1909 and 10-1918, respectively).

The sequence of the extended identifier molecule was shown below. There is a 21 nucleotide long annealing between the identifier region and the complementary identifier region. Then there is a 42 nucleotide region that represents the extended codons in the previous round of encoding. The complementary binding region that promotes the extension was a 9 nucleotide region, a 5 and 14 nucleotide region was also tested for extension. Finally there is a 14 nucleotide region that allows extension.



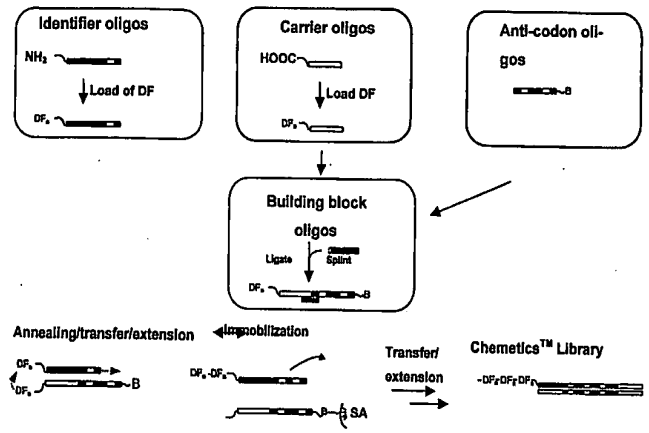
The building block oligo with a flexible linker was 5'-labeled with ^{32}P using T4 polynucleotide kinase using standard protocol (Promega, cat# 4103). This identifier molecule was annealed with the building block in the extension buffer (20 mM Hepes, 40 mM KCl, 8 mM $MgCl_2$, pH 7.4, 10 mM DTT) by heating to 80 °C for 2 min. and then slowly cooled to about 20 °C. The extension was performed using about 20 units Sequenase (USB) at 30 °C for 1 hour. The oligonucleotide complexes were then purified using micro-spin gel filtration (BioRad). Formamide dye was added to the samples before loading on a 10 % Urea polyacrylamide gel. The gel was developed using autoradiography (Kodak, BioMax film). The result of this experiment is shown on Fig. 20.

The gel shows: Lane 1, a mix of the three 5'-labeled (^{32}P) building block oligos with 5, 9 or 14 nucleotides in the binding region; lane 2, extension using a building block oligo with a 9 space linker and a 18 nucleotide binding region; lane 3, extension using a building block oligo with a 9 space linker and a 9 nucleotide binding region; lane 4, extension using a building block oligo with no linker and a 9 nucleotide binding region; lane 5, extension using a building block oligo with a 9 space linker and a 14 nucleotide binding region; lane 6, Extension using a building block oligo with a 18 space linker and a 14 nucleotide binding region; lane 7, extension using a building block oligo with no linker and a 14 nucleotide binding region; lane 8, extension using a building block oligo with no linker and a 5 nucleotide binding region.

The result shows that an efficient extension can be accomplished using a flexible linker together with a binding region. The result also shows that extension is possible without the flexible linker and only a small (5 nucleotides) binding region. The last result is an example of the loop-out setup described in the beginning of this example where the loop-out region is the 42 nucleotides described in the sequence above.

Example 7: Selection of an integrin αVβ3 ligand from a 484-member small molecule library encoded by chemetics™.

Overview of the procedure



DF : Drug fragment / functional entity
B : Biotin
SA: Streptavidin

The method for producing a library of bifunctional complexes, in which each member of the library comprises a synthetic molecule and an identifier that may be decoded to establish the synthetic history of the synthetic molecule comprises several steps, exemplified below. In a first step (General procedure 1), four different identifier oligonucleotides are loaded with a scaffold molecule or drug fragment. In this example the loading is conducted using an amino group on the identifier oligo as the attachment point for the drug fragment/ scaffold molecule. The identifiers may be regarded as the nascent bifunctional complexes.

To prepare the building block oligos, identical carrier oligos are initially loaded with eleven different drug fragments using general procedure 2. The eleven loaded carrier oligos are then ligated to anti-codon oligos of the first and the second round using general procedure 3, thereby obtaining 11 building blocks for the first round and eleven building blocks for the second round.

The library formation is described in detail in general procedure 4 and includes the mixing of the four different identifier oligos with the eleven different building blocks of the fist round. To bias the library one of the identifiers and one of the first round building blocks were added in an amount 100 below the amount of the other components. At conditions providing for annealing between the identifiers and the building blocks, a cross-link between the scaffold molecules of the identifier oligo and the drug fragments were effected. The identifier oligos were then extended using a polymerase and using the anti-codon of the building block as the identifier. After the extension, the drug fragment is released from the building block by cleavage of a linkage between the drug fragment and the oligo. The spent building block oligo is removed by streptavidin beads.

The second round includes the addition of building blocks to the nascent identifier-synthetic molecule complex obtained in the first round. To bias the library, one of the eleven second round building blocks was added in an amount 100 times below the amount used for the 10 other building blocks. The second round follows the same scheme as depicted above for the first round. The library formed is of 4 * 11 * 11 = 484 members. One of the members, which is a known ligand for the target, appears only in a concentration of the library of one out of $3 * 10^8$ bifunctional complexes.

The library is then subjected to a selection process, as disclosed in general procedure 5. The selection involves addition of the library to wells coated with immobilized target. After incubation of the library with the target, non-binding members of the library is removed by washing and a linkage between the synthetic molecule and the indentifier is cleaved. The cleaved off identifiers were collected and amplified by PCR. The amplified identifiers were decoded using general procedure 6.

General procedure 1: Loading of identifier oligos

10 µL triethanolamine (TEA) (0.1 M in DMF) was mixed with 10 µL Building Block (BB) with Pent-4-enal as an amine protection group (0.1 M in DMSO). From this mixture 6.7 µL was taken and mixed with 3.3 µL EDC [1-Ethyl-3-(3-Dimethylaminopropyl) carbodiimide Hydrochloride] (0.1 M in DMF) and incubated 30 minutes at 25°C. 10 µL of the Building block-EDC-TEA mixture was added to 10 µL of amino oligo in 0.1 M HEPES buffer ((4-(2-Hydroxyethyl)-1-

piperazineethanesulfonic acid, SIGMA), pH 7.5 and incubated with the oligo for 30 minutes.

During this half hour, another 6.7 µL of BB-TEA mix was mixed with3.3 µL EDC (0.1 M in DMF) and incubate for 30 minutes at 25°C. 10 µL of this second BB-EDC-TEA mixture was then added to the amino oligo mixture together with 10 µL of 0.1 M HEPES buffer to maintain a 1:1 ratio of DMSO/DMF: H_2O. Then the mixture was incubated for 30 minutes.

During this half hour, another 6.7 µL of BB-TEA mix was mixed with3.3 µL EDC (0.1 M in DMF) and incubate for 30 minutes at 25°C. 10 µL of this third BB-EDC-TEA mixture was then added to the amino oligo mixture together with 10 µL of 0.1 M HEPES buffer to maintain a 1:1 ratio of DMSO/DMF: H_2O. Then the mixture was incubated for 30 minutes.

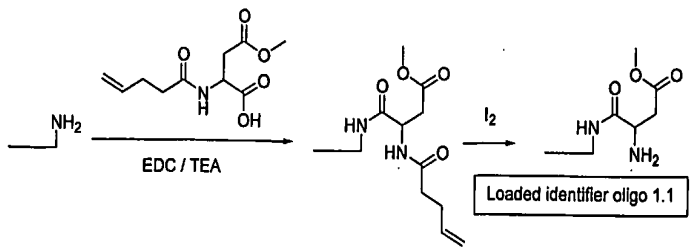
The loaded oligo was then purified by gel filtration with columns (Biospin P-6, Bio-Rad) equilibrated with water. The pent-4-enal amine protection group was then removed by addition of 0.25 volumes 25 mM I_2 in 1:1 water:tetrahydrofuran (THF) and incubation at 37°C for 2 hours. The mixture was then purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. Loaded identifier oligos were analyzed by ES-MS.

Example 7.1.1

Identifier oligo 1.1: 5'- **NSP**ACCTCAGCTGTGTATCGAGCGGCAGCGTTATCG-TCG-3'

Sequence identifying the loaded fragment

N : 5'-Amino-Modifier 5 (Glen research cat# 10-1905-90)
S : Spacer C3 CPG (Glen research cat# 20-2913-01)
P : PC Spacer Phosphoramidite (Glen research cat# 10-4913-90)

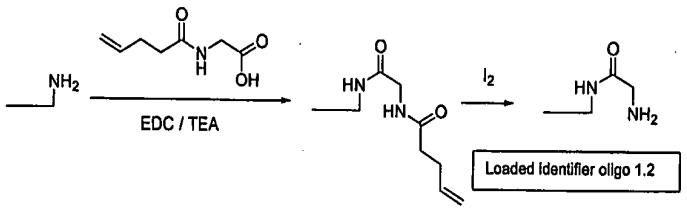


Loaded identifier oligo 1.1 analyzed by ES-MS:
Expected Mass : 11709 Da
Observed Mass : 11708 Da

Example 7.1.2

Identifier oligo 1.2: 5'- **NSP**ACCTCAGCTGTGTATCGAGCGGCAGCAGTGCCG-TCG-3'

N : 5'-Amino-Modifier 5 (Glen research cat# 10-1905-90)
S : Spacer C3 CPG (Glen research cat# 20-2913-01)
P : PC Spacer Phosphoramidite (Glen research cat# 10-4913-90)



Loaded identifier oligo 1.2 analyzed by ES-MS:
Expected Mass : 11647 Da
Observed Mass : 11641 Da

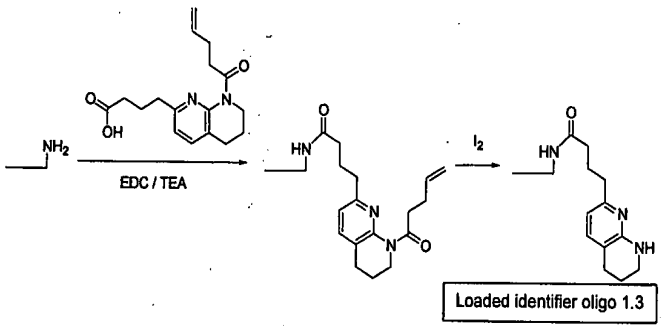
Example 7.1.3

Identifier oligo 1.3: 5'- **NSP**ACCTCAGCTGTGTATCGAGCGGCAGCGCACACG-TCG-3'

N : 5'-Amino-Modifier 5 (Glen research cat# 10-1905-90)

S : Spacer C3 CPG (Glen research cat# 20-2913-01)

P : PC Spacer Phosphoramidite (Glen research cat# 10-4913-90)



Loaded identifier oligo 1.2 analyzed by ES-MS:

Expected Mass : 11761 Da

Observed Mass : 11759 Da

Example 7.1.4

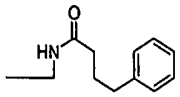
Identifier oligo 1.4: 5'- **NSP**ACCTCAGCTGTGTATCGAGCGGCAGCGGATACG-TCG-3'

N : 5'-Amino-Modifier 5 (Glen research cat# 10-1905-90)

S : Spacer C3 CPG (Glen research cat# 20-2913-01)

P : PC Spacer Phosphoramidite (Glen research cat# 10-4913-90)

Loaded identifier oligo:



Expected Mass : 11775 Da

Observed Mass : 11775 Da

General procedure 2: Loading of carrier oligo

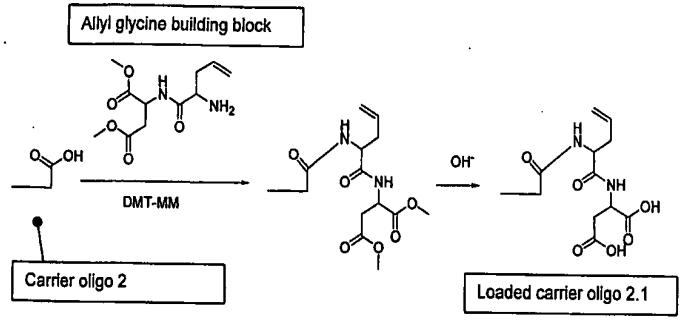
10-15 nmol of carrier oligo 2 was lyophilized and redissolved in 27.5 µl H_2O. To this was added 7.5 µl 1 M HEPES pH 7.5, 10 µl of 2-amino-pent-4-enal protected (allyl-glycine) building block (0.1 M in dimethyl sulfoxide), and 5 µl DMT-MM [4-(4,6-dimethoxy-1,3,5-thiazin-2-yl)-4-methylmorpholinium chloride] (0.5 M in water). The mixture was incubated 4-16 hours at 25-30°C. The oligo was purified by gel filtration (Biospin P-6, BioRad). To convert the methyl ester moiety of the building block to a carboxylic acid, 5 µl 0.4 M NaOH was added and the mixture was incubated 20 min at 80°C. The mixture was then neutralized by adding 10 µl 0.5 M HEPES pH 7.5 and 5 µl 0.4 M HCl. The loaded building block oligo was purified by gel filtration (Biospin P-6, BioRad) and analyzed by ES-MS

Carrier oligo 2: 3'-**2**GGAGTCGACACATAGCTCGC**p**-5'

2: Carboxy dT (Glen research cat# 10-1035-90)

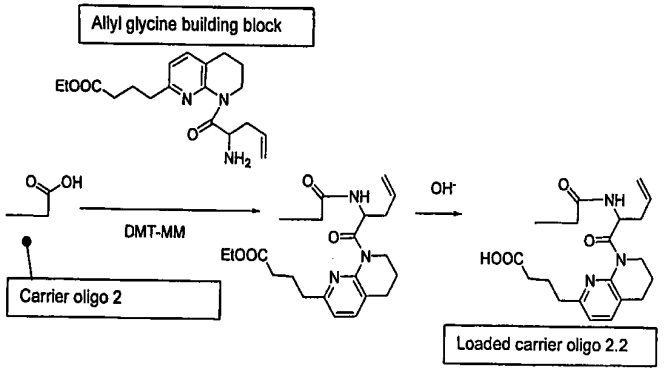
p: 5' phosphate

Example 7.2.1



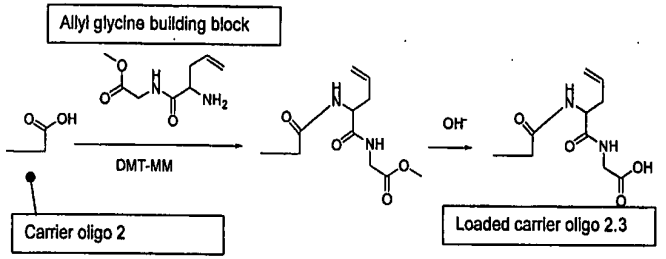
Loaded carrier oligo 2.1 analyzed by ES-MS:
Expected Mass : 6856 Da
Observed Mass : 6857 Da

Example 7.2.2



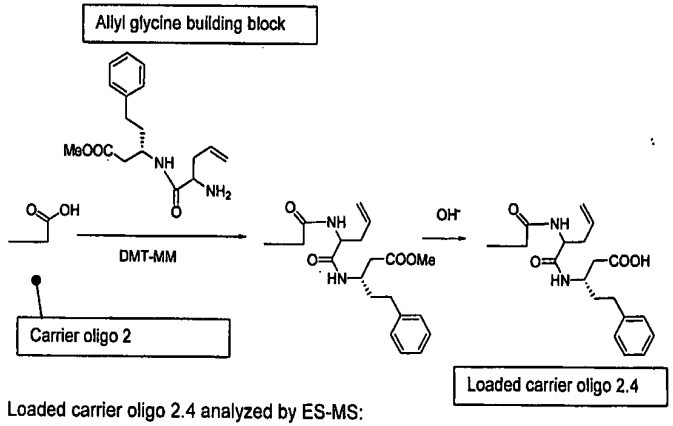
Loaded carrier oligo 2.2 analyzed by ES-MS:
Expected Mass : 6944 Da
Observed Mass : 6945 Da

Example 7.2.3



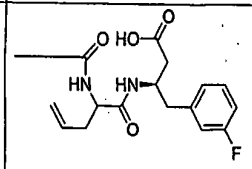
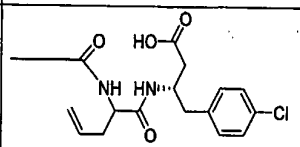
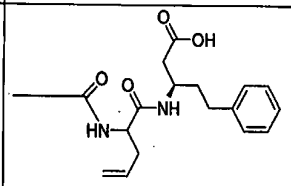
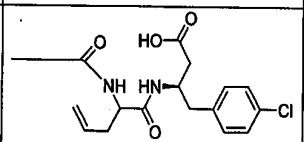
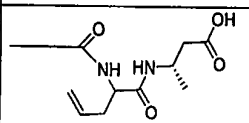
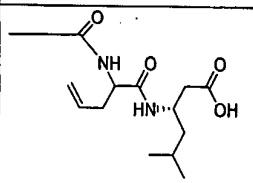
Loaded carrier oligo 2.3 analyzed by ES-MS:
Expected Mass : 6798 Da
Observed Mass : 6800 Da

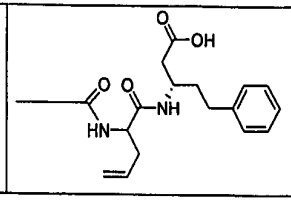
Example 7.2.4



Loaded carrier oligo 2.4 analyzed by ES-MS:
Expected Mass : 6917 Da
Observed Mass : 6919 Da

Table I

Carrier oligo Example	Structure of loaded Carrier oligo	Expected Mass	Observed Mass
7.2.5		6924	6923
7.2.6		6940	6939
7.2.7		6920	6919
7.2.8		6940	6939
7.2.9		6830	6829
7.2.10		6871	6871

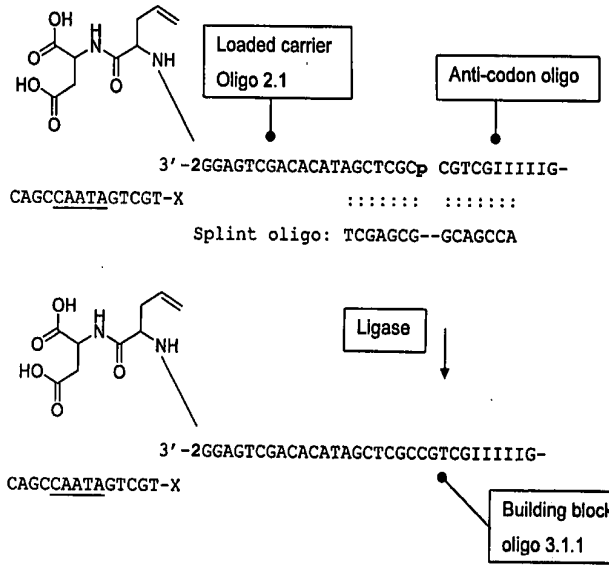
7.2.11		6920	6919

General procedure 3: ligation of anti-codon oligo with loaded carrier oligo

500 pmol loaded carrier oligo was mixed with 750 pmol anti-codon oligo and 750 pmol splint oligo. The mixture was lyophilized and redissolved in 15 µl water. Oligos were annealed by heating and slowly cooling to 20°C. 15 µl TaKaRa ligase mixture (Takara Bio Inc) was added and the reaction was incubated at 20°C for 1 hour. The mixture was purified by gel filtration (Biospin P-6, BioRad) and the efficiency of the ligation was checked by running an aliquot on a Novex TBE-UREA gel (Invitrogen).

Examples of building block oligos for first round of encoding

Example 7.3.1.1



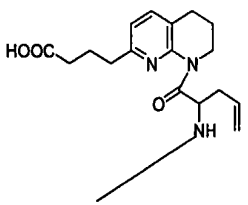
2: Carboxy dT (Glen research cat# 10-1035-90)

P: 5' phosphate

X: 5' biotin

Efficiency of ligation : > 95 %

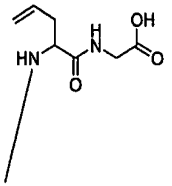
Example 7.3.1.2



3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCCGTGTGTCGT-X

Efficiency of ligation : > 95 %

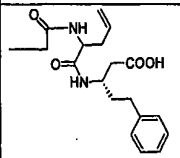
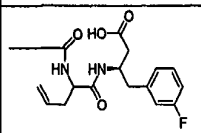
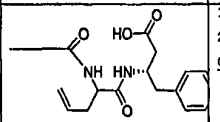
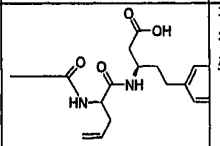
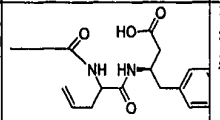
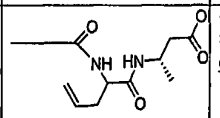
Example 7.3.1.3

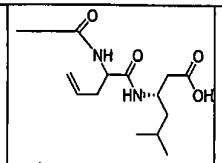
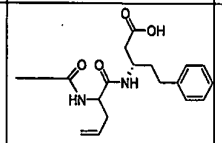


3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCTCACGGTCGT-X

Efficiency of ligation : > 95 %

Table II

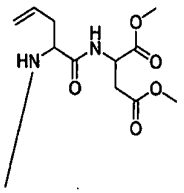
Building block oligo example	Structure of loaded Drug fragment	Building block oligo sequence 2: Carboxy dT (Glen research cat# 10-1035-90) X: 5' biotin	Ligation effi-ciency
7.3.1.4		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCCCTATGTCGT-X	> 95 %
7.3.1.5		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCGCGACGTCGT-X	> 95 %
7.3.1.6		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCGACCAGTCGT-X	> 95 %
7.3.1.7		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCACAAGGTCGT-X	> 95 %
7.3.1.8		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCTGGACGTCGT-X	> 95 %
7.3.1.9		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCGCTCGGTCGT-X	> 95 %

7.3.1.10		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCCATAGGTCGT-X	> 95 %
7.3.1.11		3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCCCGGAGTCGT-X	> 95 %

Examples of building block oligos for second round of encoding

Example 7.3.2.1

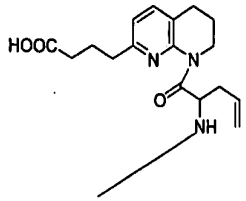
Building block oligo 3.2.1 :



3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCGTCAATA-CAGCTTAGACGGTAGATTTX

Efficiency of ligation : > 95 %

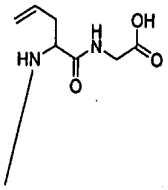
Example 7.3.2.2



3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCGTCGTGTCAG-CTTAGACGGTAGATTTX

Efficiency of ligation : > 95 %

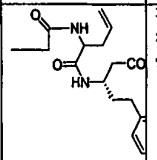
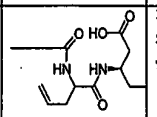
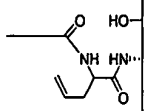
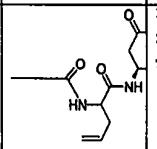
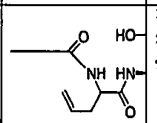
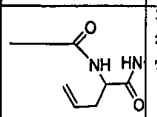
Example 7.3.2.3

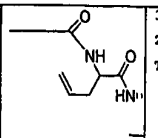
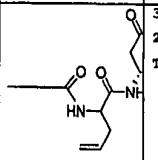


3'-2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCGTTCACG-CAGCTTAGA-CGGTAGATTTX

Efficiency of ligation : > 95 %

Table III

Building block oligo example	Structure of loaded Drug fragment	Building block oligo sequence 2: Carboxy dT (Glen research cat# 10-1035-90) X: 5' biotin	Ligation effi-ciency
7.3.2.4		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TCCTATCAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.5		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TGCGACCAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.6		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TGACCACAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.7		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TACAAGCAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.8		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TTGGACCAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.9		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TGCTCGCAGCTTAGACGGTAGATTTX	> 95 %

7.3.2.10		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TCATAGCAGCTTAGACGGTAGATTTX	> 95 %
7.3.2.11		3'- 2GGAGTCGACACATAGCTCGCCGTCGIIIIIGCAGCIIIIIGTCG TCCGGACAGCTTAGACGGTAGATTTX	> 95 %

General procedure 4: Encoding a small molecule library by chemetics™

Example 7.4.1: Encoding a 484-member small molecule library by chemetics™

Example 7.4.1.1 First encoding round

2 pmol of loaded identifier oligo 1.1 was combined with 200 pmol of each loaded identifier oligo 1.2, 1.3, and 1.4. (602 pmol loaded identifier oligos in total). These were mixed with 0.7 pmol building block oligo 3.1.3., and 72.7 pmol each of 10 different other first round building block oligos (eg. 3.1.1 and.3.1.2; 727 pmol loaded building block oligos in total). The oligos were lyophilized and redissolved in 50 µl extension buffer (EX) [20 mM HEPES, 150 mM NaCl, 8 mM $MgCl_2$]. The mixture was heated to 80°C and slowly cooled to 20°C to allow efficient annealing of identifier and building block oligos. 5 µl of 0.5 M DMT-MM in water was added and the mixture was incubated at 37°C for 4 hours.

Extension of the identifier oligo on the building block oligo identifier was performed by adding 3 µl of a 10 mM mixture of each deoxynucleotide triphosphate [dATP, dGTP, dCTP, dTTP] and 3 µl of 13 units/µl Sequenase (Amersham Biosciences). The mixture was subsequently incubated at 30°C overnight. Then 3 µl of 2M NaOH was added and the mixture was incubated for 80°C for 10 minutes followed by neutralization by addition of 3 µl 2M HCl. The mixture was then purified by passing through a gel filtration column (Biospin P-6, BioRad). 0.25 volumes of 25 mM I_2 in

1:1 THF:water was added, mixed and incubated at 37°C for 2 hours. 60 µl binding buffer (BF) [100 mM HEPES, 150 mM NaCl] and water ad 300 µl was added.

The mixture was added to streptavidin-sepharose beads (Amersham Biosciences) pre-washed 3 times in BF buffer and incubated at room temperature for 10 minutes followed by incubation on ice for 10 minutes with gentle stirring. The beads were then washed three times with water. Extended identifier oligos were stripped from the building block oligos bound to the streptaviding-sepharose beads by applying 100 µl NH3 1:1 in water and incubating at room temperature for 5 minutes.

7.4.1.2 Second encoding round

To the eluate was added 0.36 pmol second round loaded building block oligo 3.2.2 and 36.4 pmol each of 10 different other second round building block oligos (eg. 3.2.1 and 3.2.3; 364 pmol loaded second round building block oligos in total) and the mixture was lyophilized and redissolved in 50 µl EX buffer. The encoding was performed essentially as described under 7.1.1.

7.4.1.3 Final extension

The eluted identifier oligo were lyophilized and dissolved in 50 µl EX buffer. Then 200 pmol primer E38 [5'-XTTTTAGATGGCAGAT-3', X=CXS Biotin] was added. Annealing was performed by heating the mixture to 80°C and slowly cooling to 20°C. Extension of the identifier oligo was performed by adding 3 µl of a 10 mM mixture of each deoxynucleotide triphosphate [dATP, dGTP, dCTP, dTTP] and 3 µl of 13 units/µl Sequenase. The mixture was subsequently incubated at 30°C for 2 hours. The mixture was then purified by passing through a gel filtration column (Biospin P-6, BioRad). This eluated was used for selection. An aliquot (sample 7.1.3) was removed for analysis of the inpout in the selection procedure.

General procedure 5: selection

Maxisorp ELISA wells (NUNC A/S, Denmark) were coated with each 100 µL 2µg/mL integrin αVβ3 (Bachem) in PBS buffer [2.8 mM NaH_2PO_4, 7.2 mM Na_2HPO_4, 0.15 M NaCl, pH 7.2] overnight at 4°C. Then the integrin solution was substituted for 200 µl blocking buffer [TBS, 0.05% Tween 20 (Sigma P-9416), 1% bovine serum albumin (Sigma A-7030), 1 mM $MnCl_2$] which was left on for 3 hours at room temperature. Then the wells were washed 10 times with blocking buffer and the encoded library

was added to the wells after diluting it 100 times with blocking buffer. Following 2 hours incubation at room temperature the wells were washed 10 times with blocking buffer. After the final wash the wells were cleared of wash buffer and subsequently inverted and exposed to UV light at 300-350 nm for 30 seconds. Then 100 µl blocking buffer without Tween-20 was immediately added to each well, the wells were shaken for 30 seconds, and the solutions containing eluted identifiers were removed for PCR analysis (sample 5.1)

General procedure 6: analysis of selection input and output

PCR was performed on the input for (sample 7.3.1) and output of (sample 5.1) the selection using primers corresponding to the 5' end of the identifier oligos and the E38 primer. PCR was performed using Ready-To-Go (RTG) PCR beads (Amersham Biosciences) and 10 pmol each primer in a reaction volume of 25 µl. The PCR reaction consisted of an initial denaturation step of 94°C for 2 minutes followed by 30-45 cycles of denaturation at 94°C for 30 seconds, annealing at 58°C for 1 minute and extension at 72°C for 1 minute. A final extension step of 2 minutes at 72°C was included. The PCR products were resolved by agarose gel electrophoresis and the band corresponding to the expected size was cut from the gel and purified using QIAquick Gel Extraction Kit (QIAGEN).

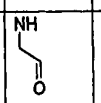
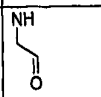
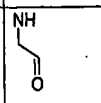
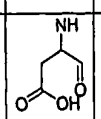
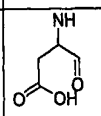
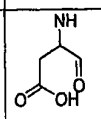
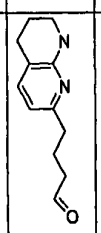
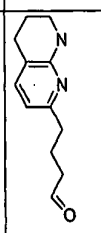
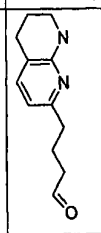
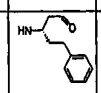
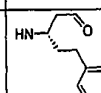
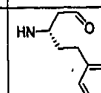
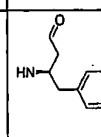
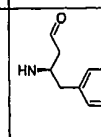
To sequence individual PCR fragments the purified PCR products were cloned into the pCR4-TOPO vector (Invitrogen) according to the manufacturer's instructions. The resulting mixture was used for transformation of TOP10 *E. coli* cells (Invitrogen) using standard procedures. The cells were plated on growth medium containing 100 µg/ml ampicillin and left at 37°C for 12-16 hours.

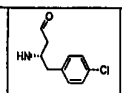
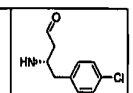
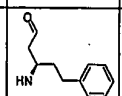
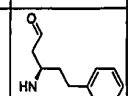
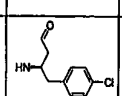
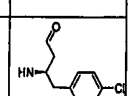
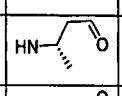
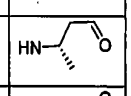
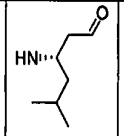
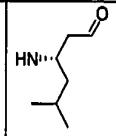
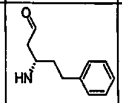
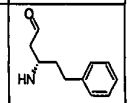
Individual *E.coli* clones were picked and transferred to PCR wells containg 50 µl water. These wells were then boiled for 5 minutes and 20 µl mixture from each well was used in a PCR reaction using RTG PCR beads and 5 pmol each of M13 forward and reverse primers according to the manufacturer's instructions. A sample of each PCR product was then treated with Exonuclease I (USB) and Shrimp Alkaline Phosphatase (USB) to remove degrade single stranded DNA and dNTPs and sequenced using the DYEnamic ET cycle sequencing kit (Amersham Biosciences) according to the manufacturer's instructions and the reactions were analyzed on a

MegaBace 4000 capillary sequencer (Amersham Biosciences). Sequence outputs were analyzed with ContigExpress software (Informax Inc.).

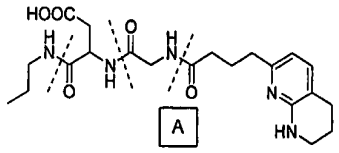
Overview of drug fragments present in the library:

Table IV

Identifier			Building block oligo for first round			Building block oligo for second round		
Oligo	Relative amount in library	Structure of drug fragment	Oligo	Relative amount in library	Structure of transferred drug fragment	Oligo	Relative amount in library	Structure of transferred drug fragment
1.1	100		3.1.1	1		3.2.1	100	
1.2	1		3.1.2	100		3.2.2	100	
1.3	100		3.1.3	100		3.2.3	1	
1.4	100		3.1.4	100		3.2.4	100	
			3.1.5	100		3.2.5	100	

			3.1.6	100		3.2.6	100	
			3.1.7	100		3.2.7	100	
			3.1.8	100		3.2.8	100	
			3.1.9	100		3.2.9	100	
			3.1.10	100		3.2.10	100	
			3.1.11	100		3.2.11	100	

The library had the potential to encode the integrin αVβ3 ligand A (Molecule **7** in Feuston B. P. et al., Journal of Medicinal Chemistry **2002**, *45*, 5640-5648) from 1 out of $3*10^8$ identifiers.



As can be seen from the table above, the library had the potential to encode ligand A for every $3*10^8$ identifiers (1 x 1 x 1 = 1 out of every 301 x 1001 x 1001 ~ $3*10^8$)

Example 7.6.1: Result of sequencing analysis of input for selection procedure and output from selection procedure.

The codon combination compatible with encoding of ligand A was not found in 28 sequences derived from the encoded library before selection in agreement with the expected low abundance of this codon combination (1 in $3*10^8$).

A codon combination compatible with encoding of ligand A was found in 5 out of 19 sequences derived from the encoded library after selection in integrin αVβ3-coated wells.
These numbers correspond to an enrichment factor of $(3*10^8 / (19 / 7)) = 8*10^7$.

Example 8: Selection of encoded molecules using size-exclusion column

This example illustrates the possibility to use column separation to perform selection on complexes against various targets. In this example, size-exclusion chromatography (SEC) is used, but other types of chromatography can be used where target-bound complexes are separated from the non-bound complexes.

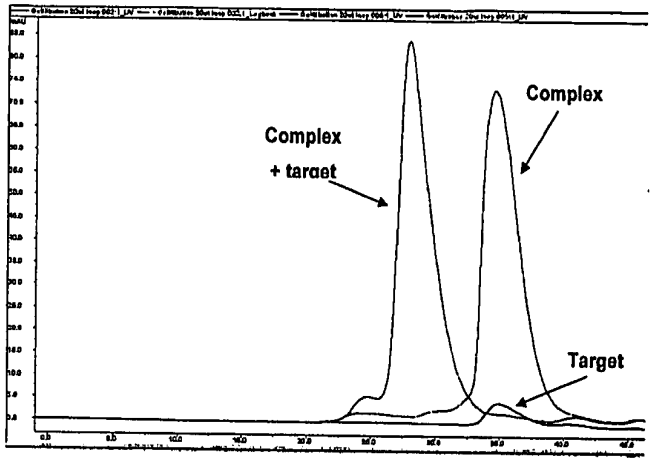
The complex is exemplified in this example by a biotin molecule attached to an oligonucleotide sequence with a predetermined sequence. Thus, the nucleotide sequence of the identifier specifies the identity of the synthetic molecule as biotin. The encoding sequence can have any length and be divided into discrete regions for encoding various building blocks as discussed elsewhere herein. Also, the displayed molecule can have a linear or scaffold structure.

Biotin-AATTCCGGAACATACTAGTCAACATGA

Biotin is known to bind to streptavidin. The binding of biotin to streptavidin will link the identifier to the target molecule and therefore change the identifiers physical and chemical properties, such as e.g. the apparent molecular weight. This change is possible to detect using e.g. size-exclusion chromatography:

78 pmol of the complex molecule was loaded on a Superdex 200, PC 3.2/30 column (ÄKTA-FPLC, AmershamPharmaciaBiotech) and analysed in PBS buffer with a flow

rate of 0.050 ml/min. As can be seen below, the complex molecules retention-time was approximately 35 minutes. When the target (83 pmol streptavidin) was analysed under identical conditions the retention-time was approximately the same. The low absorption of the target molecules is due to the wavelength (260 nm) used in the measurement. At this wavelength, the extinction coefficient is high for the nucleotides in the complexes but low for the protein target.



However, when the complex molecules was premixed with the target molecules (78 pmol complex and 83 pmol target incubated for about 1 h in PBS buffer) to allow binding and then analysed under identical conditions, the retention-time change significantly (28 minutes). The change is due to the increase in molecular weight (or hydrodynamic volume) due to the binding of the complex to the target. This will allow the separation of the target-bound complexes from the non-bound complexes. The fraction that contains the complexes and the target molecules are pooled and amplified using appropriate primers. The amplified identifiers can then be used to decode the structures of the enriched displayed molecules.

The strategy of performing column-selection of libraries of bifunctional complexes has two major advantages. First, the enriched (target-bound) complexes are eluted before the non-bound complexes, which will drastically reduce the background from

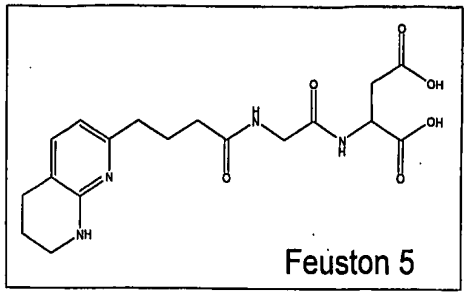
the non-bounded complexes. Secondly, the enrichment on the column will be extensive due to all the separation steps in the pores in the matrix.

The separation of the target-bound complexes using this approach will be dependent on the molecular weight of the complexes but predominantly of the molecular weight of the target. The molecular weight of the target can be adjusted by linking the target to a support that increases the apparent molecular weight. The increased molecular weight will enhance the separation by reducing the retention-time on the column. This can be done using for example a fusion protein, antibody, beads, or cross-linking the target in multimeric form. Thus, the target protein can be expressed as a fusion protein or a specific antibody can be use to increase the molecular weight. The target can be immobilized on small beads that permit separation and the target can be cross-linked using standard reagents to form multimers or cross-linked to a carrier molecule, for example another protein. Preferably, the molecular weight is increase so the target molecules elute in the void volume of the column.

Examples of other types of column separation that can be used are affinity chromatography, hydrophobic interaction chromatography (HIC), and ion-exchange chromatography. Examples of column media, other that Superdex, that can be used in size-exclusion chromatography are: Sephacryl, Sepharose or Sephadex.

Example 9: Formation of 25-member library by split-and-mix and selection of ligand

The human integrin receptor α_v/β_{III} is implicated in many biological functions such as inflammatory responses and thrombus formation as well as cellular migration and metastatic dissemination. The natural ligands for α_v/β_{III} contain an RGD tri-peptide consensus motif that interacts with the receptor binding pocket. Consequently, much medical research have focused on the synthesis and identification of small molecule RGD-mimetics with increased affinity for the α_v/β_{III} receptor. One mimetic, Feuston 5 (Feuston *et al.*, *J Med Chem.* 2002 Dec 19;45(26):5640-8.), comprising an arginine bioisostere coupled to a GD dipeptide exhibits a ten-fold increased affinity for α_v/β_{III} (K_D = 111 nM) compared to the RGD-tripeptide.



Feuston 5

Here, a 25 member small-molecule library was synthesised, comprising the Feuston 5 ligand and 24 additional small molecules a split and mix procedure. The library was screened for interaction with the receptor and the DNA was amplified by PCR, sequenced and the corresponding small-molecule ligand(s) identified.

Protocol

Library generation:

Fig. 22 and Fig. 23 show a general scheme for the synthesis of the library. Initially, a 36 nt oligo nucleotide (ID)

5'-XACCTCAGCTGTGTATCGAGCGGCAGCGGCCTCGTCG

containing a 5'-terminal amino-group (Glen Research catalog # 10-1905-90) linked by a Spacer-PEG18 (Glen Research catalog # 10-1918-90) and a photocleavable (Glen Research catalog#10-4913) spacer was synthesised by standard phosphoramidite chemistry (purchased from DNA technology A/S Denmark). 1 nmol of the ID oligonucleotide was loaded with penteneoyl-Asp(OMe)-OH using the following scaffold loading protocol A:

1 nmol ID oligonucleotide was lyophilized and then dissolved in 20 µl of 100 mM Na-borate buffer, pH 8.0 with 90 mM sulpho-N-Hydroxysuccinimide (sNHS, Merck). Pre-activation of scaffold: 15 µl of 100 mM pentenoyl-Asp(OMe)-OH in DMSO was incubated with 15 µl of 100 mM 1-ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC, Merck) in DMF and incubated for 30 min at 30°C before addition to the ID solution. Following incubation for 45 min at 30°C, additional 30 µl of pre-activated scaffold was added and the solution incubated for another 45 min at 30 °C. Excess scaffold, activation agents, solvents and salt was removed by double gel-

filtration using Bio-rad microspin columns 6 and eluted in MS-grade H_2O. Loading was verified by Electrospray-MS (Bruker Inc) analysis. Subsequently, the amino-protection group was removed by addition of 0.2 volumes of 25 mM iodine in a mixture of THF/H_2O (1:1) and incubated at 37 °C for 2 h. Excess iodine was quenched using addition of 20 mM 2-mercaptoethanol before gelfiltration purification using Bio-rad 6 microspin columns. From MS-analysis the loaded and deprotected ID oligonucleotide was estimated to be > 75 % pure (data not shown).

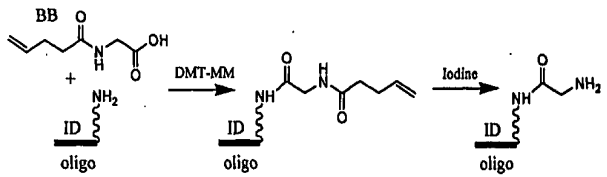
500 pmol of D-loaded ID oligo was annealed to 500 pmol complementary oligo with the sequence 5'-TGTGCGACGAGGCCGCTGC
by denaturation for 2 min at 80 °C followed by slow cooling to ambient temperature. The double stranded oligo pair (ID-ds) with a 4 nt overhang (for efficient annealing and ligation) was used in a split & mix reaction protocol shown schematically below using the following procedure:

Addition of position 2 codons and free reactants: 500 pmol of ID-ds was split into 5 wells (here, eppendorf tubes). 100 pmol of a specific 2nd position codon oligonucleotides of the sequence

Z1-0: pCACAAGTACGAACGTGCATCAGAG/pTCCTCTCTGATGCACGTTCGTACT
Z1-1: pCACATAGTCTCCTCCACTTCCATG/pTCCTCATGGAAGTGGAGGAGACTA
Z1-2: pCACATACATCGTTCCAGATACCG/pTCCTCATGGAAGTGGAGGAGACTA
Z1-3: pCACATCCAGTGCAAGACTGAACAG/pTCCTCTGTTCAGTCTTGCACTGGA
Z1-4: pCACAAGCATCACTACTCTGTCTGG/pTCCTCCAGACAGAGTAGTGATGCT

was added to each well and the oligos ligated in a volume of 20 µl using ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP] and 10 units T4-DNA ligase at ambient temperature for 1 hour.

Subsequently, the 5 ligation products were purified individually using Biorad 6 spin columns according to manufacturer's instructions and lyophilized. Next, a specific reactant was reacted with the scaffold according to the scheme shown in Fig. 22 using loading protocol A described above. Excess free reactant, reagents and buffer was removed by gelfiltration. The elute was pooled, lyophilized and resuspended in 40 µl of H_2O before addition of 10 µl of 25 mM iodine (in THF/H_2O, ratio 1:1) for deprotection.



Reaction of N-penteneoyl protected glycin reactant with an ID oligo and subsequent deprotection using Iodine. The reaction was incubated at 37°C for 2 h. Excess Iodine was quenched by addition of 1 µl of 1 M 2-mercaptoethanol and left at ambient temperature for 5 min before purification of the sample using spin-gelfiltration (Biorad 6).

The sample was split into 5 wells for addition of 3rd position codons using the codon oligonucleotides:

Z0-0: pAGGACGAGCAGGACCTGGAACCTGGTGCGTTCCTCCACCACGTCTCCG/
pGCACCAGGTTCCAGGTCCTGCTCG
Z0-1: pAGGACTCGACCACTGCAGGTGGAGCTCCGTTCCTCCACCACGTCTCCG/
pGGAGCTCCACCTGCAGTGGTCGAG
Z0-2: pAGGACGTGCTTCCTCTGCTGCACCACCGGTTCCTCCACCACGTCTCCG/
pCGGTGGTGCAGCAGAGGAAGCACG
Z0-3: pAGGACCTGGTGTCGAGGTGAGCAGCAGCGTTCCTCCACCACGTCTCCG/
pGCTGCTGCTCACCTCGACACCAGG
Z0-4: pAGGACTCGACGAGGTCCATCCTGGTCGCGTTCCTCCACCACGTCTCCG/
pGCGACCAGGATGGACCTCGTCGAG

p = 5' phosphate.

and reacted with free reactant as described for the 2nd position and shown on Fig. 22 with the following exception: The F3 reactant did not react efficiently using protocol A due to poor solubility of F3 in organic solvent. Consequently, F3 was reacted using the following procedure (protocol B): The ligated and lyophilized sample was

dissolved in 35 µl 100 mM Na-borate buffer (pH 8.0) before addition of 10 µl 100 mM F3 reactant in water and 5 µl of 500 mM 4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4 methylmorpholinium chloride (DMT-MM, carboxylic acid activator) and incubated at 25 °C for 2 h. Following the coupling reaction, excess reactant, reagent and salt was removed by gelfiltration as described in protocol A. The remaining steps were conducted as described for position 2.

Prior to conducting the selection step, a strand exchange reaction was performed in order to assure that no mis-annealed oligos was assembled. The strand-exchange was done by annealing of 200 pmol of AH361 oligo (5'-CGGAGACGTGGTGGAGGAAC-3') in sequenase buffer containing 200 µM deoxyribonucleotides (dNTP) in a total volume of 80 µl before addition of 20 units of sequenase and incubation at 30 °C for 1 h. Following extension the reaction mixture was used in the selection step without further purification.

Selection

Maxisorp ELISA wells (NUNC A/S, Denmark) were coated with each 100 µL 2µg/mL integrin αVβ3 (Bachem) in PBS buffer [2.8 mM NaH_2PO_4, 7.2 mM Na_2HPO_4, 0.15 M NaCl, pH 7.2] overnight at 4°C. Then the integrin solution was substituted for 200 µl blocking buffer [TBS, 0.05% Tween 20 (Sigma P-9416), 1% bovine serum albumin (Sigma A-7030), 1 mM $MnCl_2$] which was left on for 1 hour at room temperature. Then the wells were washed 2 times with 250 µl blocking buffer and 5 µl of the encoded library was added to the wells after diluting it 20 times with blocking buffer. Following 2 hours incubation at room temperature the wells were washed with 20 x 250 µl blocking buffer.

ELUTION

After the final wash the wells were cleared of wash buffer and subsequently inverted and exposed to UV light at 300-350 nm for 30 seconds using a trans-illuminator set at 70% power.

100 µl blocking buffer without Tween-20 was immediately added to each well, the wells were shaken for 30 seconds, and the solutions containing eluted templates were removed for PCR analysis.

PCR amplification

PCR on input and output use primers corresponding to the 5' end of Frw-27 oligo (ACCTCAGCTGTGTATCGAG) and the AH361 primer. 5 µl eluted DNA was used for PCR in a 25 µl reaction using 10µl Eppendorph hotmastermix 2.5x and 10 pmol each of AH361 & Frw-27. PCR was run: (ENRICH30): 94°C 2 min, then 30 cycles of [94°C 30 sec, 58°C 1 min, 72°C 1 min], then 72°C 10 min.

Cloning and Sequencing

The TOPO-TA (Invitrogen Cat#K4575-J10) ligation was reacted with 4 µl PCR product, 1 µl salt solution, 1 µl vector. The reaction was incubated at RT for 30 min. Heat-shock competent TOP10 E.coli cells was thawed and put on ice. 5 µl ligation reaction was added. Following 30 min on ice, the cells were heat-shocked at 42°C water for 30 sec, then put on ice. 250 µl SOC was added and the cells incubated 1 h at 37°C, before spreading on LB-ampicillin plates followed by incubation ON at 37°C.

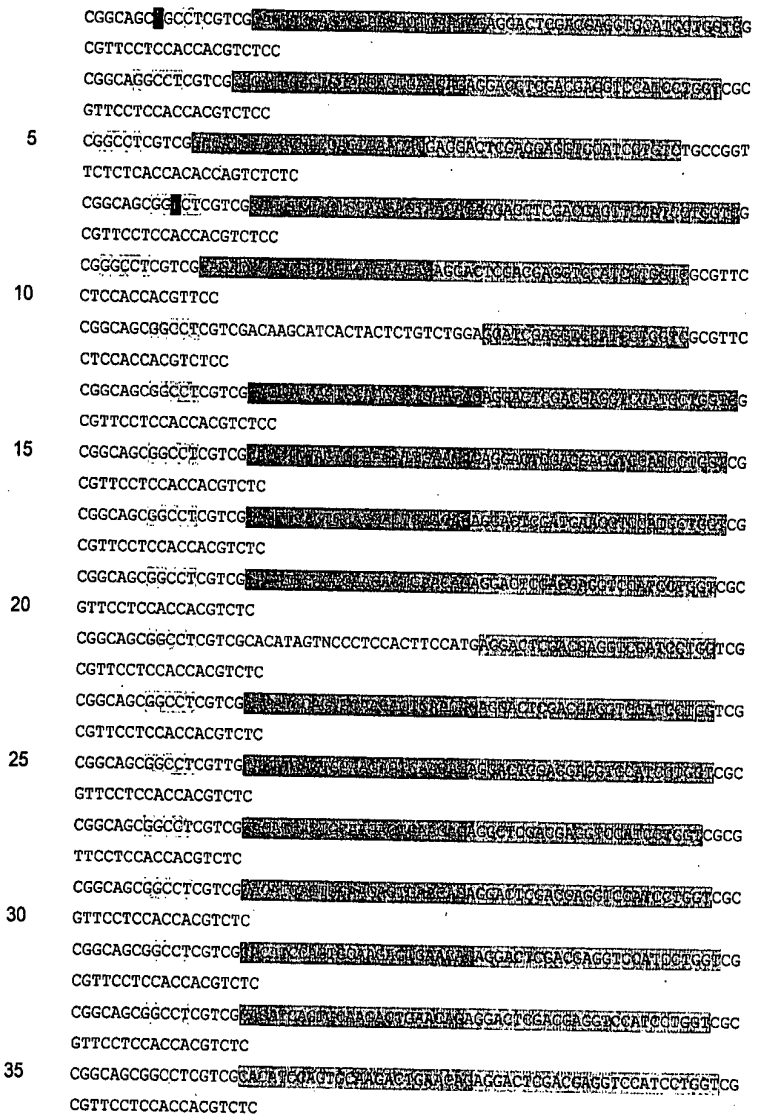
Individual *E.coli* clones were picked and transferred to PCR wells containing 50 µl water. Colonies were incubated at 94°C for 5 minutes and 20 µl was used in a 25 µl PCR reaction with 5 pmol of each TOPO primer M13 forward & M13 reverse (AH365/AH366) and Ready-To-Go PCR beads (Amersham) using PCR program EKO50: 94°C 2 min, then 30 x (94°C 4 sec, 50°C 30 sec, 72°C 1 min) then 72°C 10 min.

Primers and free nucleotides were degraded by adding 1 µl EXO/SAP mixture 1:1 to 2 µl PCR product. Incubation was at 37oC for 15 min and then 80oC for 15 min. 5 pmol T7 primer (AH368) was added and water to 12 µl. subsequently, 8 µl DYEnamic ET cycle sequencing Terminator Mix was added followed by PCR-cycling using 30 rounds of (95°C 20 sec, 50°C 15 sec, 60°C 1 min). Purification was done using seq96 spinplates (Amersham), followed by analysis on a MegaBace sequenizer.

Library sequence output

18 successful sequences were informative of the isolated DNA from the selection step and are shown below

Output sequences



5-mer sequences highlighted at position 1 corresponds to aspartic acid, the highlighted 20-mer sequence at position 2 (central) corresponds to glycin and the highlighted 20-mer sequence (+4 bases from ligation overhang) corresponds to the F3 building block. Thus, 16 out of 18 sequences identify the exact Feuston-5 ligand (F3-G-D) as the single dominant small molecule that bind the integrin αv/βIII receptor. Note that only the F3 BB is identified in position 3 arguing for very strong bias towards this arginine bioisostere.

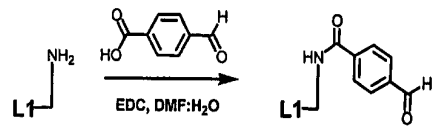
The data shows that chemical synthesis of small molecule library, tagging, selection and identification procedure is highly efficient using this technology which is expectedly easily scalable and applicable to libraries comprising more than 10^8-10^{10} different molecules.

Fig. 22 shows an overview of the library generation using a unique 1st position oligo loaded with D (aspartate), 5 different reactant/oligo pairs in the 2nd position and 5 different reactant/oligo pairs in the 3rd position. Ultimately, a library of 1x5x5 = 25 different trimers each attached to their corresponding unique DNA code is assembled. Arrowheads indicate site of ligation.

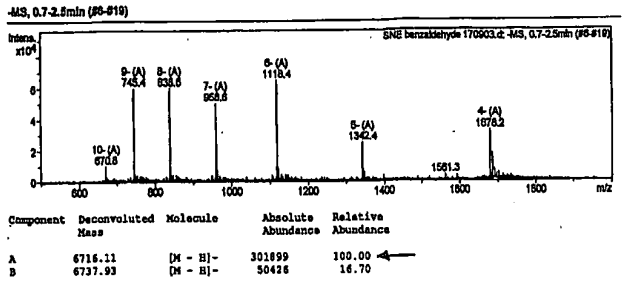
Example 9: Encoded Multi Component Reaction (MCR) product

9.1 Preparation of aldehyde-comprising scaffold-oligo, using 4-carboxybenzaldehyde

A solution of 4-carboxybenzaldehyde (scaffold) in DMF (25 µL, 150 mM) was mixed with 25 µL of a 150 mM solution of EDC in DMF. The mixture was left for 30 min at 25°C. 50 µL aminooligo (10 nmol) in 100 mM HEPES buffer pH 7.5 was added and the reaction mixture was left for 20 min at 25°C. Excess scaffold was removed by extraction with EtOAc (500 µL) and remaining EtOAc was removed *in vacuo* by spinning 10 min in a speedvac. The mixture was then purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. The loaded oligo were analyzed by ES-MS.



Mw 6585 Mw 6716



Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	6716.11	[M - H]-	301899	100.00
B	6737.93	[M - H]-	50426	16.70

Aminooligo **L1,1** used in section 9.2: Mw = 7154

L1.1: 5'-5CG ATG GTA CGT CCA GGT CGC AX

3'X = 3'Biotin

5'5 = 5' amino C6 (Glen Research catalogue # 10-1906-90)

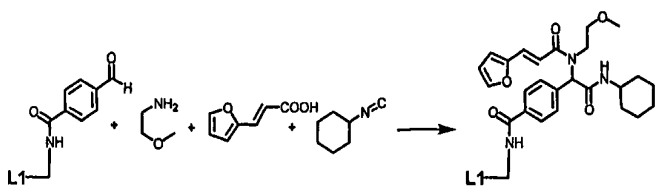
Aminooligo **L1.2** used in section 9.3: Mw = 6585

L1.2: 5'-GCG ACC TGG AGC ATC CAT CGX

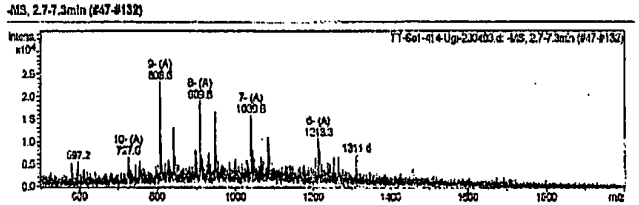
3'X = Amino-C2-dT-3'-PO_4 (Glen Research catalogue # 10-1037-90)

9.2 Multi-component reaction

A solution of Benzaldehyde loaded **L1.1** oligo (200 pmol) was lyophilized and redissolved in 10µL H_2O. 2-Methoxy ethylamine in methanol (10µL, 40mM), 3-furan-2-yl-acrylic acid in methanol (10µL, 40 mM), and cyclohexyl isocyanide in methanol (10µL, 40mM) was added and incubated overnight at 37 °C. The reaction mixture was diluted with 40µL H_2O and purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. MCR-product on oligo was analyzed by ES-MS. The starting benzaldehyde loaded **L1** oligo (**A**) was identified in the MS-spectrum together with the UGI product (**B**).



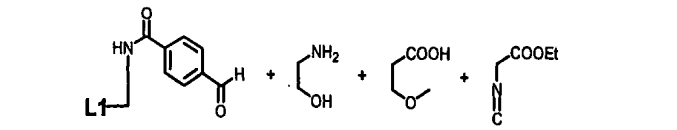
A: Mw 7285 **B:** Mw 7589

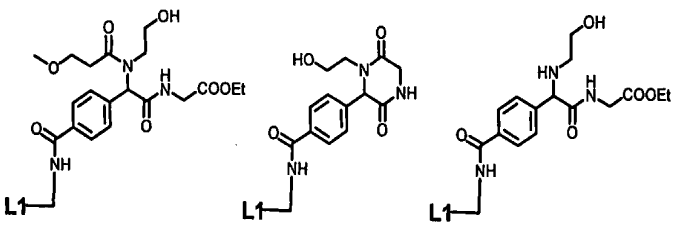


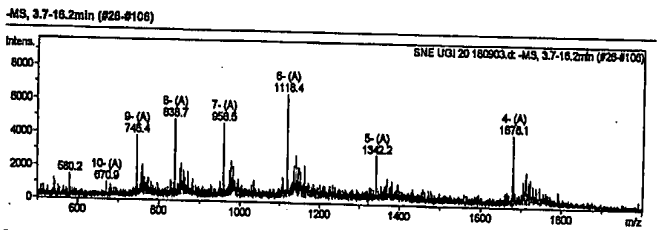
Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	7205.25	[M - H]-	75430	100.00
B	7589.64	[M - H]-	52950	70.20
C	7307.54	[M - H]-	30548	40.50
D	7463.73	[M - H]-	25475	33.77

9.3 Multi-component reaction

A solution of benzaldehyde loaded **L1.2** oligo (320 pmol) was lyophilized and redissolved in 10µL H_2O. 2-Amino ethanol in methanol (10µL, 40mM), 3-Methoxy-propionic acid in methanol (10µL, 40 mM), and ethyl isocyanoacetate in methanol (10µL, 40mM) was added and incubated overnight at 37 °C. The reaction mixture was diluted with 40µL H_2O and purified by gel filtration with spin columns (Biospin P-6, BioRad) equilibrated with water. MCR-product on oligo was analyzed by ES-MS. The starting benzaldehyde loaded **L1** oligo (**A**) was identified in the MS-spectrum together with three products, **B** Diketopiperazine, **C** UGI product and **H** the Amine product.







Component	Deconvoluted Mass	Molecule	Absolute Abundance	Relative Abundance
A	6716.10	[M - H]-	27984	100.00
B	6845.98	[M - H]-	13154	47.01
C	6976.15	[M - H]-	8913	31.85
D	6818.39	[M - H]-	10156	36.29
E	6887.71	[M - H]-	10187	36.40
F	6872.09	[M - H]-	8871	31.70
G	6913.59	[M - H]-	5551	19.84
H	6894.50	[M - H]-	4579	16.36

9.4 Encoding

Excess reactants, activation agents, solvents and salt was removed by double gel-filtration using Bio-rad microspin columns 6 and eluted in MS-grade H_2O and loading

was verified by Electrospray-MS (Bruker Inc) analysis before the displayed molecule attached to the oligonucleotide L1 was encoded.

The benzaldehyde loaded oligonucleotide L1.1 in section 9.2, that has been reacted with the other three components to form the displayed molecule as describe above was mixed with the codon oligonucleotides L2, L3 and L4 together with the splint oligonucleotides S1, S2 and S3 (sequences shown below) and ligated using a ligase (T4 DNA ligase). The ligation was performed using the following conditions. The double stranded oligonucleotide was achieved by mixing the encoding strands (L1, L2, L3 and L4) with the splint oligonucleotides (S1, S2 and S3) to form a 7 oligonucleotide hybridisation product (for efficient annealing and ligation). About 50 pmol of each specific oligonucleotide was used and the oligonucleotides was ligated in a volume of 20 µl using ligation buffer [30 mM Tris-HCl (pH 7.9), 10 mM MgCl2, 10 mM DTT, 1 mM ATP] and 10 units T4-DNA ligase at ambient temperature for 1 hour.

L1: 5'-CGATGGTACGTCCAGGTCGCA-3'
S1: 5'-ATCGTGCTGCGACCT-3'
L2: 5'-GCACGATATGTACGATACACTGA-3'
S2: 5'-GTGCCATTCAGTGT-3'
L3: 5'-ATGGCACTTAATGGTTGTAATGC-3'
S3: 5'-TGTATGCGCATTAC-3'
L4: 5'-GCATACAAATCGATAATGCAC-3'

The identifier comprising the tags was amplified using a forward (FP) and reverse (RP) primer using the following conditions: 5 µl of the ligated indentifier oligonucleotide was used for PCR in a 25 µl reaction using 10µl Eppendorph hotmastermix 2.5x and 10 pmol each of AH361 & Frw-27. PCR was run: (ENRICH30): 94°C 2 min, then 30 cycles of [94°C 30 sec, 58°C 1 min, 72°C 1 min], then 72°C 10 min.

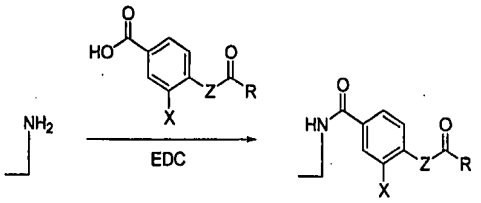
FP: 5'-CGATGGTACGTCCAGGTCGCA-3'
RP: 5'-GTGCATTATCGATTTGTATGC-3'

The amplified identifier oligonucleotide was cloned to verify that the assembled oligonucleotides contained the codon region (CGTCC, GTACG, AATGG and TCGAT).

The TOPO-TA (Invitrogen Cat#K4575-J10) ligation was reacted with 4 µl PCR product, 1 µl salt solution, 1 µl vector. The reaction was incubated at RT for 30 min. Heat-shock competent TOP10 E.coli cells was thawed and put on ice. 5 µl ligation reaction was added. Following 30 min on ice, the cells were heat-shocked at 42°C water for 30 sec, and then put on ice. 250 µl SOC was added and the cells incubated 1 h at 37°C, before spreading on LB-ampicillin plates followed by incubation ON at 37°C.

Individual E.coli clones were picked and transferred to PCR wells containing 50 µl water. Colonies were incubated at 94°C for 5 minutes and 20 µl was used in a 25 µl PCR reaction with 5 pmol of each TOPO primer M13 forward & M13 reverse (AH365/AH368) and Ready-To-Go PCR beads (Amersham) using PCR program: 94°C 2 min, then 30 x (94°C 4 sec, 50°C 30 sec, 72°C 1 min) then 72°C 10 min. Primers and free nucleotides were degraded by adding 1 µl EXO/SAP mixture 1:1 to 2 µl PCR product. Incubation was at 37°C for 15 min and then 80°C for 15 min. 5 pmol T7 primer (AH368) was added and water to 12 µl. Subsequently, 8 µl DYEnamic ET cycle sequencing Terminator Mix was added followed by PCR-cycling using 30 rounds of (95°C 20 sec, 50°C 15 sec, 60°C 1 min). Purification was done using seq96 spinplates (Amersham), followed by analysis on a MegaBace sequenizer.

Example 10: Loading of entity onto tag.



Loaded

Procedure:

25 µL of a 150 mM building block solution in DMF was mixed with 25 µL of a 150 mM solution of EDC in DMF. The mixture was left for 30 min at 25°C. 50 µL of an aminooligo (10 nmol) in 100 mM HEPES buffer pH 7.5 was added and the reaction

mixture was left for 20 min at 25°C. The excess building block was removed by extraction with EtOAc (500 µL). The excess EtOAC was removed at reduced pressure in a speedvac. The building block loaded aminooligo was ethanol precipitated twice using NH4OAc and analysed by electron spray mass spectrometry (ES-MS).

Example 11:

The following example illustrates the use of the tagging principle for the identification of entities comprising desirable properties isolated from a library of entities. The principle is shown schematically in figure 1.

DNA-tagging of peptides for the identification of complexes that bind the integrin receptor αV/β3.

Materials:

- Purified human integrin αV/β3 (Chemicon Inc.)
- Streptavidin Sepharose 6B (AmershamPharmacia)
- Nunc ImmunomoduleU8 Maxisorp (Biotecline cat# Nun-475078)
- Sheared herring DNA (Sigma)
- Taq-polymerase (Promega) and 10 X Taq-pol buffer
- Binding buffer [100 mM NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5]
- UV-transilluminator
- SPDP [*N*-succinimidyl 3(2-pyridyldithio)propionate] (Molecular Probes, Cat: S-1531)
- Micro Bio-Spin 6 (Bio-Rad cat: 732-6221)
- 6 tagging oligo nucleotides with the following sequences:

TO#1: 5'-XCTATGCGGACTGACTGGTAC-3'
TO#2: 5'-XCTATGATGCTTAGGCGGTAC-3'
TO#3: 5'-XCTATGTACCGTACGTGGTAC-3'
TO#4: 5'-XCTATGAATGCTAGCTGGTAC-3'
TO#5: 5'-XCTATGGATTGCGCGTGGTAC-3'
TO#6: 6'-XCTATGCCACTATTAGGGTAC-3'

where X = 5' C6 amino modifier (Glen research cat# 10-1916-90) suitable for attachment of functional entities such as peptides, small molecules or polymers.

- Complementary (Template) oligonucleotides with the following sequences:

CO#1:
5'-BPTATA**GGATCC**GTACCAGTCAGTCCGCATAG**GAATTC**TAGT-3'
CO#2:
5'-BPTATA**GGATCC**GTACCGCCTAAGCATCATAG**GAATTC**TAGT-3'
CO#3:
5'-BPTATA**GGATCC**GTACCACGTACGGTACATAG**GAATTC**TAGT-3'
CO#4:
5'-BPTATA**GGATCC**GTACCAGCTAGCATTCATAG**GAATTC**TAGT-3'
CO#5:
5'-BPTATA**GGATCC**GTACCACGCGCAATCCATAG**GAATTC**TAGT-3'
CO#6:
5'-BPTATA**GGATCC**GTACCCTAATAGTGGCATAG**GAATTC**TAGT-3'

Where, B = 5'-biotin (Glen research Cat#10-1953-95) and P = photocleavable linker (Glen research cat#10-4913-90).
The underlined 10 nucleotide sequences are unique for each tagging oligonucleotide and have a unique complementary oligonucleotide counterpart.
Sequences highlighted in bold are suitable for cloning purposes.

- Oligonucleotides for PCR amplification

AO#1: 5'-BPTATAGGATCCGTACC-3'
AO#2: 5'-ACTAGAATTCCTATG-3'

- 6 peptides with the following composition

P#1: GRGDSPC
P#2: GRADSPC
P#3: GRGESPC
P#4: GDGRSPC

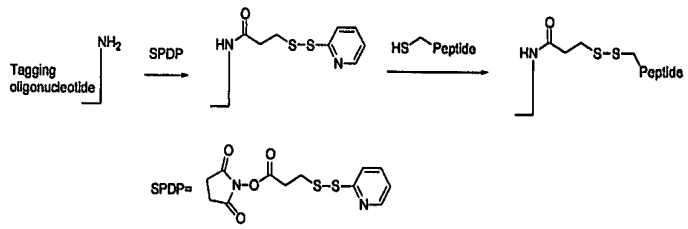
P#5: CKKK
P#6: CFFKKK

A = Alanine, G = Glycin, R = Arginine, D = Aspartate, P = Proline, F = Phenylalanine, K = Lysine and E = Glutamate. All peptides are end-capped by N-terminal carboxylation and C-terminal amidation. Peptides were supplied by Schafer-N A/S, DK-Denmark.

Protocol

Step1: Tagging of peptides #1-6 with a specific oligonucleotide (TO#1-6).

Each TO oligonucleotide contains a single 5'end amino nucleophile (X) which can be covalently linked to the cysteine thiol-group of a peptide using the heterobifunctional cross-linker SPDP in the following reaction.



Procedure: 5 nmol amino-oligo is dried and resusspended in 160 µl of 100 mM Hepes-OH, (pH 7.5). 40 µl 20 mM SPDP (in DMSO) is added and incubate for 2 h at 30°C. The sample is extracted with 3 x 500 µl ethylacetate and dried for 10 min in a speedvac. The sample is purified using microbio-spin 6 column equilibrated with 100 mM Hepes-OH. Add 10 µl of.1 M peptide and incubate at 25 °C for 2 h. Precipitate twice with 2 M NH_4OAc/Ethanol. Redissolve in 50 µl H_2O and verify tagging by Electrospray-MS analysis (Bruker Inc.).

Step 2: Anneal complementary oligonucleotides (CO#1-6) to TO-peptide complexes from step 1.

Procedure:

10 pmol of TO#1-6 loaded with their corresponding peptide is added to a mixture comprising 20 pmol each of CO#1-6 in binding buffer [100 mM NaCl, 5 mM $MgCl_2$, 50 mM Hepes-OH, pH 7.5] and a total volume of 100 µl. The sample is heated to 80 °C for 2 minutes and slowly cooled to room temperature over 30 minutes.

Step 3: Purify doublestranded DNA-peptide complexes (Optionall).

Following annealing, only tagged molecules that have annealed to their complementary oligonucleotide sequences will comprise both a functional entity and a biotin handle (see figure 1). Consequently, to reduce "noise" in the selection step, single-stranded tagged-molecules can be removed from the library in a pre-selection step using the biotin handle.

Procedure:

50 µl Streptavidine-sepharose 6B Slurry is washed in 3 X 1 ml binding buffer before resuspending the beads in 100 µl binding buffer. The CO/TO-peptide annealing mixture is added to the straptavidine beads and incubated at 25 °C for 30 min with agitation. Subsequently, the streptavidine beads are pelleted, the supernatant is discarded and the beads are washed three times with 1 ml of binding buffer. The beads are resuspended in 100 µl binding buffer binding buffer and finally, the CO/TO-peptide complexes are released using photocleavage. The photocleavage reaction is conducted by incubating the sample on a Vilber-Lourmat UV-transilluminator TFX-20.M for 30 seconds at 70% effect. The eluted CO/TO-peptide complexes are removed to a new tube.

Step 4: Enrich library for ligands that bind the integrin αV/β3 receptor.

The library of molecules is tested for binding to the integrin αV/β3 receptor immobilised on a plastic surface.

Procedure:

A single well of a Nunc 8 plate is incubated overnight with 100 µl of 1 µg/ml of integrin receptor in standard phosphate-buffered saline (PBS). The well is washed five

times with 100 µl PBS followed by blocking using 100 µl 0.5mg/ml sheared herring DNA in PBS-buffer for 2 h at room temperature.
Finally the well is washed five times using 100 µl Integrin binding buffer [Tris-HCl (pH 7.5), 137 mM NaCl, 1 mM KCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$ and 1 mM $MnCl_2$].

The CO/TO-peptide complexes are added to the immobilised integrin and incubated at 37 °C for 30 min. The supernatant is removed and the immobilised integrin is washed 5 times using 100 µl integrin binding buffer. The CO/TO-ligand complexes are eluted heating the sample to 80 °C for 5 min. The sample is cooled to room-temperature. 1 µl of the sample is used for PCR amplification using 10 pmol each of AO#1 and 2 as external primers in a reaction comprising 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1 mM $MgCl_2$, 0.1 % Triton X-100, 250 mM each of dATP, dCTP, dGTP and dTTP. The sample is run with initial denaturation at 94 °C, for 2 min and 30 cycles using denaturation at 94°C for 30 seconds, annealing at 44 °C for 30 seconds and elongation at 72°C for 15 seconds. Finally, the sample is precipitated

Step 5: Isolate single stranded templates.

For subsequent selection and amplification rounds the non-template strand of the amplified PCR products should be should be removed. This step is conducted using specific purification of the biotinylated template oligo.

Procedure:

50 µl of streptavidine-sepharose 6B is washed three times with 1 ml of binding buffer. The washed beads are incubated with 25 µl (<10 pmol) of PCR product from step 4 in 100 µl binding buffer for 30 min at 25 °C. Spin the sample briefly to collect beads. Remove supernatant and wash five times using 800 µl H_2O. The beads are resuspended in 500 µl 10 mM NaOH for 2 min at room temperature. The supernatant is removed and the beads are resuspended in 100 mM biotin in 100 µl H_2O. For elution the sample is incubated at 95 °C for 10 min with agitation. Subsequently, the excess biotin is removed by Micro-spin gel-filtration.

Step 6: Anneal the new population of template oligos to the library of tagged peptides from step1.

The new population of single stranded template oligonucleotides which are enriched for sequences that represent ligands for the integrin αV/β3 receptor are annealed to the library of tagged-peptides from step1 as described in step 2 and subjected to yet another round of selection and amplification.

The selection and amplification procedure (step2-6) is repeated for 5 rounds.

Step 7: Identification of ligands.

The identity of enriched double stranded DNA fragments specific for a ligand entity or entities is established by DNA cloning in a M13mp18 plasmid vector and examining individual clones by sequence analysis. For statistical purposes more than 30 clones is sequenced to identify dominant sequence(es) within the pool of cloned sequence tags. Since the dominant DNA sequence cloned corresponds to a ligand the sequence bias directly identifies the ligand candidate(s) suitable for further examination.

Example 12

The following example illustrates the use of the tagging principle for the identification of a DNA sequence representing a small molecule isolated from a library of sequences. The principle is shown schematically in the figures.

DNA-tagging of biotin and glutathione for the identification of complexes that bind streptavidine.

Materials:

- Streptavidin Sepharose 6B (AmershamPharmacia)
- Taq-polymerase (Promega) and 10 X Taq-pol buffer
- Binding buffer [100 mM NaCl, 5 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5]
- SPDP [*N*-succinimidyl 3(2-pyridyldithio)propionate] (Molecular Probes, Cat: S-1531)
- N-hydroxysuccinimidylester-biotin (Fluka#14405)
- Glutathione (Sigma)
- Micro Bio-Spin 6 (Bio-Rad cat: 732-6221)
- T7 Exonuclease (gene 6) and 5 x buffer

- Tagging oligo nucleotides with the following sequences:

TO#1: 5'-XCTATGCGGACTGACTGGTAC-3'
TO#2: 5'-XCTATGANNNNNNNNCGGTAC-3', (65.536 sequence combinations)

where X = 5' C6 amino modifier (Glen research cat# 10-1039-90) suitable for attachment of functional entities such as peptides, small molecules or polymers. N is G, A, T or C

- Complementary (Template) oligo nucleotides with the following sequences:

CO#1:
5'-$T_sA_sT_s$**AGGATCC**GTACCAGTCAGTCCGCATAG**GAATTC**TAGT-3'
CO#2:
5'-$T_sA_sT_s$**AGGATCC**GTACCGNNNNNNNNTCATAG**GAATTC**TAGT-3'

Where, S denotes the position of a phosphorothioate in the DNA backbone.
The underlined 10 nucleotide sequences are unique for each tagging oligonucleotide or pool of tagging oligonucleotides and have a unique complementary oligonucleotide counterpart. Sequences highlighted in bold are suitable for cloning purposes.

- Oligonucleotides for PCR amplification

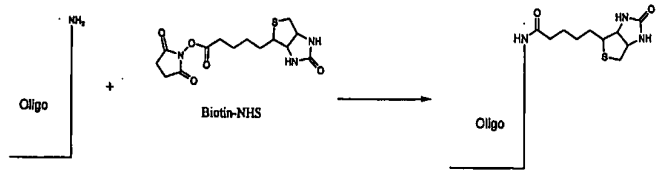
AO#1: 5'-$T_sA_sT_s$AGGATCCGTACC-3'
AO#2: 5'-ACTAGAATTCCTATG-3'

Where, S denotes the position of a phosphothioate in the DNA backbone.

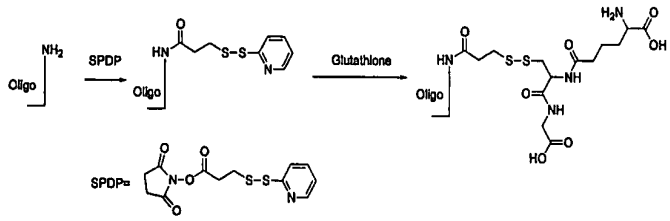
Protocol

Step1: Tagging biotin with TO#1 and tagging glutathione with TO#2.

All TO oligonucleotides contain a single 5'end amino nucleophile(X) which can be used for covalent linking of small molecules. Biotin is linked to the TO#1 amino-group using NHS-biotin (Merck) in the following reaction.



Glutathione is linked to the pool of oligonucelotides using the heterobifunctional cross-linker SPDP in the following reaction.



Procedure:

Tagging of biotin with TO#1:

5 nmol of TO#1 oligonucleotide is dried down and resuspended in 80 µl 100 mM Hepes-OH buffer (pH 7.5). 20 µl of 50 mM NHS-Biotin (in DMSO) is added to the oligonucleotide and the sample incubated at 30 oC for 2 hours. The sample is extracted twice using 200 µl ethyl-acetate before purification on a Micro-spin 6 column. Tagging of biotin is verified using Electrospray-MS (Bruker Inc.).

Tagging of glutathione (GSH) with TO#2:

5nmol of TO#2 is dried down and resusspended in 160 µl of 100 mM Hepes-OH, (pH 7.5). 40 µl 20 mM SPDP (in DMSO) is added and the sample is incubated for 2 h at 30°C. The sample is extracted with 3 x 500 µl ethylacetate and dried for 10 min in the speedvac. The sample is purified using microbio-spin 6 column equilibrated

with 100 mM Hepes-OH. 10 µl of 0.1 M GSH is added and the sample is incubated at 25 °C for 2 h. Precipitate twice with 2 M NH_4OAc/Ethanol. Redissolve in 50 µl H_2O and verify tagging by Electrospray-MS analysis (Bruker Inc.).

The single biotin sequence tag and the 65.536 different glutathione sequence tags comprise a total of 65.537 different sequence-tags. The library is mixed to comprise equi-molar amounts of each sequence tag. Consequently, the library consists of 65.536-fold excess of tagged glutathione over tagged biotin.

Step 2: Anneal complementary oligonucleotides (CO#1 & 2) to TO complexes from step 1.

Procedure:

A total of 10 pmol of tagged library molecules is added to a mixture comprising 20 pmol of template molecules (CO#1 & 2) comprising 65.536 fold excess of CO#2 over CO#1 in a binding buffer [100 mM NaCl, 5 mM $MgCl_2$, 50 mM Hepes-OH, pH 7.5] and a total volume of 100 µl. The sample is heated to 80 °C for 2 minutes and slowly cooled to room temperature over 30 minutes.

Step 3: Purify doublestranded DNA complexes (Optionall).

Following annealing, only tagged molecules that have annealed to their complementary oligonucleotide sequences will comprise both a functional entity and a phosphorothioate backbone handle (see figure 1). Consequently, to reduce "noise" in the selection step, single-stranded tagged-molecules can be removed from the library in a pre-selection step using the phosphorothioate handle.

Procedure:

50 µl of activated thiopropyl-sepharose slurry is washed in 3 X 1 ml binding buffer before resuspending the beads in 100 µl binding buffer. The CO/TO annealing mixture is added to the thiopropyl-sepharose beads and incubated at 30 °C for 30 min with agitation. Subsequently, the beads is pelleted, the supernatant discarded and the beads is washed three times with 1 ml of binding buffer. The beads is resuspended in 100 µl binding buffer binding buffer and finally, the CO/TO complexes are

released using by incubation with 100 µl of 50 mM DTT in binding buffer. The eluted CO/TO complexes are removed to a new tube.

Step 4: Enrich library for ligands that binds to streptavidine.

The library of molecules is tested for binding to the streptavidine sepharose 6B.

Procedure:

50 µl of streptavidine-sepharose 6B slurry is washed three times with 1 ml of binding buffer. 10 µl of library molecules eluted at step 3 is incubated with the streptavidine in 100 µl of binding buffer for 10 minutes at 25 °C with agitation. Subsequently, the sample is washed five times using 1 ml of binding buffer. The ligand DNA is eluted by incubating of the sample in 100 µl H2O at 95 oC for 5 minutes. The sample is cooled to room-temperature. 1 µl of the sample is used for PCR amplification using 10 pmol each of AO#1 and 2 as external primers in a reaction comprising 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1 mM $MgCl_2$, 0.1 % Triton X-100, 250 mM each of dATP, dCTP, dGTP and dTTP. The sample is run with initial denaturation at 94 °C, for 2 min and 30 cycles using denaturation at 94°C for 30 seconds, annealing at 44 °C for 30 seconds and elongation at 72°C for 15 seconds. Finally, the sample is precipitated

Step 5: Isolate single stranded templates.

For subsequent selection and amplification rounds the non-template strand of the amplified PCR products should be should be removed. This step is conducted using specific purification of the template oligo strand comprising a phosphorothioate backbone.

Procedure:

The doublestranded PCR product is subjected to exonuclease digestion using phage T7 (gene 6) exonuclease. This enzyme is a doublestrand specific 5' exonuclease that is inhibited by the presence of phosphorothioate in the DNA backbone. 20 µl of doublestranded PCR product from step 4 is incubated in exonuclease T7 buffer before addition of 50 units of T7 exonuclease enzyme. The sample is incubated at 30 °C for 10 minutes. The sample is extracted once with 100 µl phenol before precipitation using NH_4-acetate/ethanol. Resuspend sample in H_2O.

Step 6: Anneal the new population of template oligos to the library of tagged molecules from step1.

The new population of single-stranded template oligonucleotides which are enriched for sequences that represent ligands for the streptavidine is annealed to the library of tagged molecules from step1 as described in step 2 and subjected to yet another round of selection and amplification.
The selection and amplification procedure (step2-6) is repeated for 5 rounds.

Step 7: Identification of ligands.

The identity of enriched double stranded DNA fragments specific for a ligand entity or entities is established by DNA cloning in a M13mp18 plasmid vector and examining individual clones by sequence analysis.
For statistical purposes more than 30 clones is sequenced to identify dominant sequence(es) within the pool of cloned sequence tags. Since the dominant DNA sequence cloned corresponds to a ligand the sequence bias directly identifies the ligand candidate suitable for further examination.

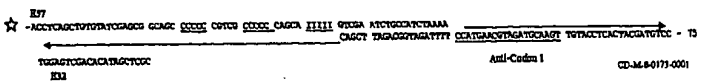
Example 13: Encoding onto an identifier obtained from a pool-encoding procedure (Mode 1) using separated compartment encoding procedure (Mode 2).

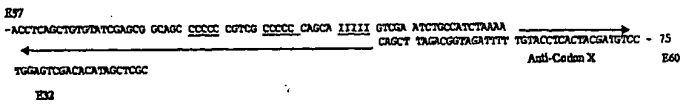
This example describes the experimental conditions used to perform Mode 2 encoding of reactants on an identifier that contains codons that have been obtained using Mode 1 encoding. The mode 1 encoding is performed as described in previous example, notably example 7. The example illustrates the general principle of combining encoding Mode 1 and 2.

Extension of the encoded identifier and transfer of the reactant is performed in separate wells where one specific zipper building block and one specific anti-codon oligonucleotide that codes for the functional entity loaded to for the zipper building block is mixed. This approach can also be used for free reactants.

Extension using the encoding Mode 2.

In this example, a radioactive labelled identifier oligonucleotide (E57) is mixed with a specific zipper building block (E32) and an anti-codon oligonucleotide (CD-M-8-0172-0001) with the anti-codon sequence (Anti-codon 1) as shown below. In another experiment, a different anti-codon oligonucleotide (E60) with a different anti-codon sequence (Anti-Codon X) was used as a reference sample.





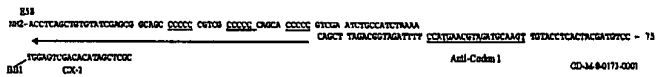
The oligonucleotide combinations (as shown below) were mixed together in separate compartments to allow specific annealing of the pairs of zipper building block and anti-codon oligonucleotides. The extension was performed in an extension buffer (20 mM Hepes, 8 mM MgCl, 150 mM NaCl) using 1 pmol identifier oligonucleotide, 2 pmol zipper building block, 2 pmol anti-codon in a final volume of 10 µl. The oligonucleotides were heated and then allow re-annealing slowly from 80 – 20°C in a PCR-machine. After annealing a mix sample (~20 ul) of 0.5 mM dNTP and 13 U Sequenase was added and the extension was run for 1 h 30°C. The sample was then analyzed by 10% urea polyacrylamide gel electrophoresis as shown on Fig. 31A.

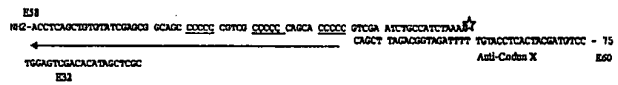
The gel analysis shows that the identifier is completely extended both with the long anti-codon 1 (lane 2) and the shorter anti-codon X (lane 3). The result also shows that there is no scrambling between the anti-codon oligonucleotides if they first are allowed to anneal to the identifier oligonucleotide before they are mixed and extended (lane 4).

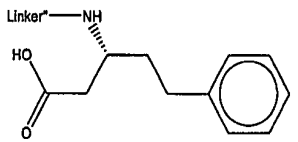
Cross-linking.

The Mode 2 encoding of a reagent on a Mode 1 encoded molecule was tested using an identifier with three codons and a displayed molecule. The transfer of the reactants is illustrated in this example by a cross-linking of a reactant on a zipper building block to the displayed molecule in the identifier oligonucleotide. The transfers were tested with cross-linking to simplify the analysis on the gel but are not limited to this type of reaction.

The functional entity as shown below (chemical structure) is linked to the oligo to form the zipper building block (CX-1) that anneals to the identifier oligonucleotide through the complementary region. This zipper building block was used together with an anti-codon oligonucleotide (CD-M-8-0172-0001) with the anti-codon sequence (Anti-codon 1) as shown below. In another (control) experiment, a different zipper building block (E32) was used together with a different anti-codon oligonucleotide (E60) with a different anti-codon sequence (Anti-Codon X).







The oligonucleotide combinations were mixed together in separate compartments to allow specific annealing of the pairs of zipper building block and anti-codon oligonu-

cleotides. The annealing to place in a extension buffer (20 mM Hepes, 8 mM MgCl, 150 mM NaCl) using 1 pmol Identifier oligonucleotide, 2 pmol zipper building block, 2 pmol anti-codon in a final volume of 10 µl. The oligonucleotides were heated and then allow re-annealing slowly from 80 – 20°C in a PCR-machine. Cross-linking was performed by adding 5 mM DMT-MM reagent and incubation for 2 h 37°C. The sample was then analyzed by 10% urea polyacrylamide gel electrophoresis as shown in Fig. 31 B.

The gel analysis shows that the functional entity on the zipper building block (CX-1) is cross-linked to the identifier oligonucleotide which contains the codons (lane 2) while the zipper building block lacking the reagent (E32) is unable to react with the identifier oligonucleotide. The result also shows that there is no scrambling between the zipper building block oligonucleotides if they first are allowed to anneal to the identifier oligonucleotide before they are mixed and cross-linked (lane 4).

While the invention has been described with reference to specific methods and embodiments, it will be appreciated that various modifications and changes may be made without departing from the invention.

Example 14: Enrichment of coding parts from bifunctional complexes displaying integrin receptor αvβ3 ligands.

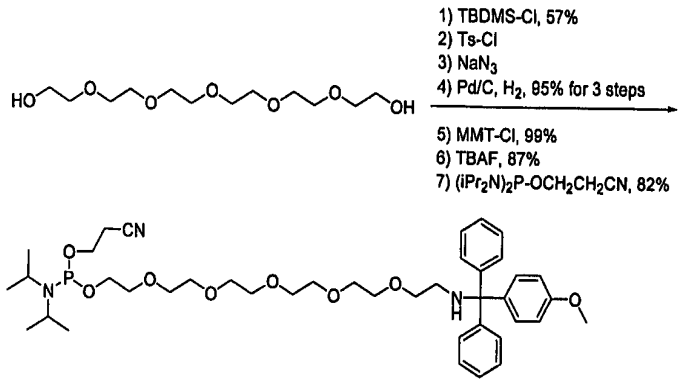
Example 14.1 Formation of L1-RGD and L1-cRGD:

An oligonucleotide L1-NH_2 was used to tag cRGD peptide and RGD peptide respectively by mixing 4 nano mol of L1-NH2 in 80 µL mixture of 500 mM Hepes KOH, pH 7.5 and 20 µL 25 mM BMPS in DMSO and then incubating for 2 hours at 30°C. The BMPS activated L1-NH2 oligonucleotide was washed three times with EtOAc, to remove unbound BMPS, and excess EtOAc was evaporated off by vacuum distillation. Ten µL 100 mM of cRGD or RGD peptide respectively was added and the reaction was incubated at room temperature over night. After incubation the cRGD or RGD tagged oligonucleotide was cleared of excess unbound peptide by gelfiltration. Tagging of peptide was confirmed by mass spectrometry analysis and the tagged products are referred to as L1-cRGD and L1-RGD respectively.

Materials:

L1-NH2 (6-8-CAGCTTGGACACCACGTCATAC, 6=LH193, 8=PCspacer) was aquired from DNA-Technology, Aarhus, DENMARK, PCspacer is a photo cleavable spacer (Glen Research Products cat# 10-4913), BMPS (*N*-[ß-Maleimidopropyloxy]succinimide ester) was acquired from Pierce (cat# 22298).

LH193 (Diisopropyl-phosphoramidous acid 2-cyano-ethyl ester 2-[2-(2-[2-[2-(2-{[(4-methoxy-phenyl)-diphenyl-methyl]-amino}-ethoxy)-ethoxy]-ethoxy}-ethoxy)-ethoxy]-ethyl ester) was synthesized according to the following method:



Commercially available hexaethylene glycol was selectively mono TBDMS protected using TBDMS-Cl, imidazole and DMAP in 57% yield. The free alcohol function was converted to an amine function by a standard three step protocol: Activation with tosyl chloride in pyridine followed by nucleophilic displacement with sodium azide in DMF and subsequent reduction using Pd/C and H2 in overall 95% for the three steps. The amine function was then protected with 4-monomethoxy trityl (MMT) in 99% yield and the TBDMS group removed using TBAF. The free hydroxyl function was finally reacted with cyanoethyl-N,N,N',N'-tetraisopropylphosphorodiamidite under tetrazole catalysis and gave the desired compound in 82% yield. ^{31}P nmr ($CDCl_3$) = 148.5 ppm.

14.2 Formation of L1-F5:

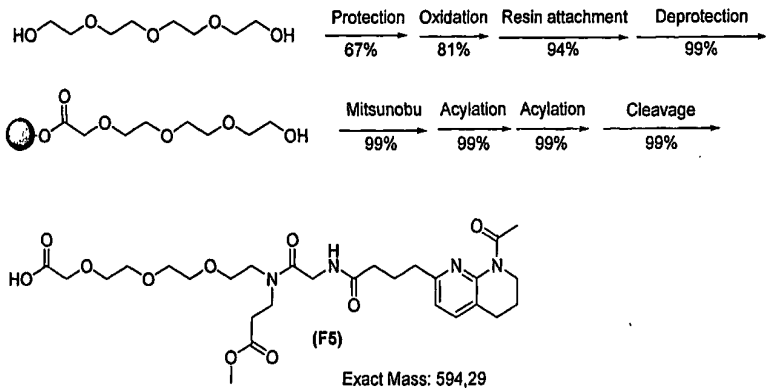
A mixture of 5 nano mol L1-NH2, 50 mM DMTMM and 10 mM F5 in 100 mM Na-Borat pH 8.0 in 50 µL was incubated over night at 30°C and the oligonucleotide tagged F5 was cleared of excess unbound F5 by gelfiltration. Subsequently the

loaded oligo was dried down by speed vacuum distillation and resuspended in 100 mM NaOH in 50 micro L and left over night at 50°C for deprotection (ester cleavage and N-acetamide cleavage) of the F5 molecule. After deprotection the suspension was neutralised with HCl and loading of F5 was confirmed by mass spectrometry analysis. Tagging of F5 was confirmed by mass spectrometry analysis. The tagged product is referred to as L1-F5.

Materials:

DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) was prepared from commercially available *N*-methylmorpholine and 2-chloro-4,6-dimethoxy-1,3,5-triazine according to Kaminski et. al (JOC (1998), 63, 4248-55).

F5 (ester and *N*-acetyl derivative) was synthesized according to the following method:



Exact Mass: 594,29

The tagged product L1-F5 carries a free carboxylic acid (the homoglycine moiety) and a free amino group (the cyclic aminopyridine).

The commercially available tetraethylene glycol was mono protected with 4-nitrobenzoic acid using 4-nitrobenzoyl chloride in 67% yield. A subsequent oxidation of the remaining primary alcohol to the corresponding carboxylic acid was performed in 81% yield using a mixture of TEMPO, chlorite and hypochlorite. The compound was attached to a 2-chlorotrityl chloride resin and subsequently treated with KCN in

MeOH-DMF (1:5) to deprotect the 4-nitrobenzoyl ester. N-(2-nitrophenyl sulfonyl) activated beta-alanine methyl ester was attached using a Mitsunobu protocol. The 4-nitrophenylsulfonyl group was removed by treatment with excess mercaptoethanol/DBU and the formed amine then acylated with two subsequent building blocks using standard Fmoc-chemistry. The final molecule was cleaved from the resin by treatment with 0.4M HCl in ether-DCM (1:4). $[M+H]^+$ (calc.) = 595.30. $[M+H]^+$ (found) = 595.28.

14.3 Formation of library component 1.

Formation of double stranded tagged L1-cRGD (referred to as L1-cRGD-T1):
Two hundred pico mol of a DNA template, T1
(GCTAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGATATC*ACCACATCTC TAGCAGCTA*GTATGACGTGGTGTCCAAGCTG)
was annealed to 50 pico mol of L1-cRGD. Subsequently the L1-cRGD oligo was extended by DNA polymerase (Sequenase).
Sequenase was acquired from Upstate Biotechnology (Cat# 70775Y).

14.4 Formation of library component 2.

Double stranded tagged L1-RGD. Referred to as L1-RGD-T2:
Two hundred pmol T2
(GCTAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGATATC*AGGTCTTCT GTCTTCTTCC*GTATGACGTGGTGTCCAAGCTG)
was annealed to 50 pico mol of L1-RGD. Subsequently the L1-RGD oligo was extended by DNA polymerase as described above.

14.5 Fomation of library component 3.

Double stranded L1-F5. Referred to as L1-F5-T3:
Two hundred pmol T3
(GCTAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGATATC*TTCAGTTCTC GACTCCTGA*GTATGACGTGGTGTCCAAGCTG)
was annealed to 50 pico mol of L1-F5. Subsequently the L1-F5 oligo was extended by DNA polymerase as described above.

14.6 Formation of library component 4.

Double stranded L1-NH2. Referred to as L1-NH2-T4:

Fifty pmol T4
(GCTAGAGACGTGGTGGAGGAAGTCTTCCTAGAAGCTGGATAT*CTGACGTGTT GACGTACACA*GTATGACGTGGTGTCCAAGCTG)
was annealed to 200 pico mol of L1-NH2. Subsequently the L1-NH2 oligo was extended by DNA polymerase as described above.

A total of 50 pico mol of each of the library components L1-cRGD-T1, L1-RGD-T2, L1-F5-T3 and L1-NH2-T4 was produced.

14.7 Enrichment

Enrichment of integrin binding complexes was performed by coating 0.04 µg/well integrin receptor αvβ3 in Nunc Immunomodule U8 Maxisorp wells (Biotecline cat # nun-47507).

In one experiment (Fig 32), L1-cRGD-T1, L1-RGD-T2, L1-F5-T3 and L1-NH2-T4 where mixed in ratios 1 pmol of L1-NH2-T4 complex, 1/1000000 pmol of L1-cRGD-T1 complex, 1/100000 pmol of L1-RGD-T2 complex and 1/10000 pmol of L1-F5-T3 complex in 100 µL buffer A (Tris buffered saline, 0.05% Tween 20, 1% Bovine serum albumin, 0.1 mg/mL herring sperm DNA). Incubation in integrin coated wells was done for 90 min at 25°C. After ligand binding all wells were washed 20 times with 250 µL buffer A during one hour. Thereafter 100 µL buffer A was applied to each well and the wells where exposed to UV light at 350 nano meters for 30 seconds in order to cleave the PC spacer thereby releasing the DNA templates from the ligand molecule. Following exposure to UV light the elution volume was removed immediately and analysed for the presence of DNA strands T1, T2, T3 and T4 by quantitative polymerase chain reaction (Q-PCR).

In a similar experiment (Fig 33), L1-cRGD-T1, L1-RGD-T2, L1-F5-T3 and L1-NH2-T4 where mixed in ratios 1 pmol of L1-NH2-T4 complex, 1/10000 pmol of L1-cRGD-T1 complex, 1/10000 pmol of L1-RGD-T2 complex and 1/10000 pmol of L1-F5-T3 complex in 100 µL buffer A. Otherwise assay conditions where as described above.

For 5 mL premix (for one 96-well plate) 2.5 mL Taqman Universal PCR Master Mix (Applied Biosystems) was mixed with 450 µL RPv2 (GTCAGAGACGTGGTGGAG-

GAA) (10 pmol/µl), 25 µL Taqman probe (6-FAM-TCCAGCTTCTAGGAAGAC-MGBNFQ; 50 µM) and 1075 µL H_2O

40.5 µL premix was aliquoted into each well and 4.5 µL of relevant upstream PCR primer (FPv2 (CAGCTTGGACACCACGTCATAC) (for standard curve) or one of the template specific primers P1 (GTCATACTAGCTGCTAGAGATGTGGTGATA) specific for T1, P-2 (CATACGGAAGAAGACAGAAGACCTGATA) specific for T2, P-3 (TCATACTCAGGAGTCGAGAACTGAAGATA) specific for T3 or P-4 (CATACTGTGTACGTCAACACGTCAGATA) specific for T4; 10 pmol/µL) and 5 µL sample (H_2O in wells for negative controls) was added.

The samples for the standard curve were prepared by diluting T4 to 10^6 copies/5 µL and subsequently performing a 10-fold serial dilution of this sample. 5 µL was used for each Q-PCR reaction.

Thermocycling/measurement of fluoresence was performed on an Applied Biosystems ABI Prism 7900HT real-time instrument utilizing the cycling parameters: 95°C 10 min, 40 cycles of 95°C 15 sec, 64°C 1 min.

From figure 32 it can be seen that the double stranded DNA complexs L1-cRGD-T1, L1-RGD-T2, L1-F5-T3, when considering input ratio compared to enriched output ratio, are enriched approximately 1 million fold, 100000 fold and 30000 fold respectively over the L1-NH2-T4 complex. L1-cRGD-T1, L1-RGD-T2 and L1-F5-T3 could not be detected after incubation in wells without integrin receptor.

Figure 33 shows enrichment of L1-cRGD-T1, L1-RGD-T2 and L1-F5-T3 respectively. The ligand DNA complexes are enriched differently. This is most likely due to different dissociation constants for the three molecules. L1-cRGD-T1, L1-RGD-T2 and L1-F5-T3 could not be detected after incubation in wells without integrin receptor.

Claims

1. A method for obtaining a bifunctional complex comprising a display molecule part and a coding part, wherein a nascent bifunctional complex comprising a chemical reaction site and a priming site for enzymatic addition of a tag is reacted at the chemical reaction site with one or more reactants, and provided with respective tag(s) identifying the reactant(s) at the priming site using one or more enzymes.

2. The method according to claim 1, wherein the bifunctional complex obtained in claim 1, is reacted further one or more times with a reactant and is provided with respective identifying tags to produce a reaction product as one part of the bifunctional complex and a coding part comprising tags which code for the identity of the reactants which have participated in the formation of the reaction product.

3. The method according to claims 1, 2 or 3, wherein the tag is a nucleic acid sequence.

4. The method according to claims 1 to 3, wherein the one or more tags is/are provided at the priming site of the nascent bifunctional complexes utilizing an enzymatic extension reaction.

5. The method according to claim 1 to 4, wherein a linking moiety is connecting the chemical reaction site and the priming site.

6. The method according to claim 5, wherein the linking moiety is a nucleic acid sequence of a length suitable for hybridisation with a complementing oligonucleotide.

7. The method according to claims 5 or 6, wherein the linking moiety is attached to the chemical reaction site via a spacer comprising a selectively cleavable linker.

8. The method according to claims 1 or 2, wherein the chemical reaction site is a chemical scaffold comprising a chemical structure having on or more reactive groups or precursors of such reactive groups.

9. The method according to any of the claims 1 to 8, wherein two or more reactants are involved in the reaction with chemical reaction site and the priming site prior to or subsequent to the reaction is provide with respective tags identifying the reactants.

10. The method according to any of the claims 1 to 9, wherein the priming site comprises a 3'-OH or 5'-phosphate group of a nucleotide, or functional derivatives thereof.

11. The method according to any of the claims 1 to 10, wherein the addition of the tag is performed by an enzyme selected from polymerase, ligase, kinase, recombinase, or a combination thereof.

12. The method according to claim 1, wherein the addition of the tag is performed prior to the reaction of the reactant with the chemical reaction site.

13. The method according to any of the claims 1 to 12, wherein the nucleic acid part of the bifunctional complex is double stranded when reacting the reactant with the chemical reaction site.

14. The method according to any of the preceding claims, wherein the reactant is a free reactant.

15. The method of any of the claim 1 to 13, wherein the reactant comprises a nucleic acid sequence and a functional entity capable of being reacted at the chemical reaction site.

16. The method according to claim 15, wherein the reactant is a building block comprising an oligo nucleotide sufficiently complementary to the linking molety to allow for hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity.

17. The method according to any of the claims 1 to 16, wherein the tag comprises a codon identifying the reactant and an anti-tag comprises a nucleic acid sequence complementary to the codon (anti-codon).

18. The method according to claim 17, wherein the tag further comprises a framing sequence informative of the point in time of the synthesis history the associated reactant has reacted.

19. The method according to claim 1, wherein the tag comprises an oligonucleotide of 3, 5, 8, 11, 15, 20, 30 or more nucleotides.

20. The method according to any of the preceding claims, wherein the bifunctional complex comprises codons of different lengths.

21. The method according to any of the claims 1 to 20, wherein an oligonucleotide comprising a sequence complementing the tag oligonucleotide is immobilized.

22. The method according to claim 1, wherein, instead of using enzymatic addition of a tag, the tag is chemically connected to the priming site applying a guiding oligonucleotide complementing an end of the tag and a part of the bifunctional complex comprising the priming site, such that the ends abut each other.

23. A method for obtaining a bifunctional complex comprising a display molecule part and a coding part, comprising the steps of

a) providing a nascent bifunctional complex comprising a reactive group and an oligonucleotide identifier region,

b) providing a building block comprising an oligonucleotide sufficient complementary to the identifier region to allow for hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity,

c) mixing nascent bifunctional complex and building block under hybridisation conditions to form a hybridisation product,

d) transferring the functional entity of the building block to the nascent bifunctional complex through a reaction involving the reactive group of the nascent bifunctional complex, and

e) enzymatically extending the oligonucleotide identifier region to obtain a codon attached to the bifunctional complex having received the functional entity.

24. The method according to claim 23, further comprising steps f) separating the components of the hybridisation product and recovering the complex.

25. The method according to claim 23 and 24, said display molecule part comprising the reaction product of functional entities and the reactive group of the initial complex, wherein steps c) to f) are repeated as appropriate.

26. The method according to any of claim 23 to 25, wherein the reactive group of the nascent bifunctional complex is part of a chemical scaffold.

27. The method according to any of claims 23 to 26, wherein the enzyme for production of the codon is a polymerase.

28. The method according to any of the claims 23 to 27, wherein the anticodon is unique.

29. The method and the claims 23 to 28, wherein a similar combination of nucleotides of an anti-codon does not appear on another building block carrying another functional entity.

30. A method for generating a library of bifunctional complexes, comprising the steps of:

a) providing one or more different nascent bifunctional complexes comprising a reactive group and an oligonucleotide identifier region,

b) providing a plurality of different building blocks, each comprising an oligonucleotide sufficient complementary to an identifier region to allow for

hybridisation, a transferable functional entity, and an anti-codon identifying the functional entity,

c) mixing nascent bifunctional complexes and plurality of building blocks under hybridisation conditions to form hybridisation products,

d) transferring functional entities of the building blocks to the nascent bifunctional complexes through a reaction involving the reactive group of the nascent bifunctional complex,

e) enzymatically extending the oligonucleotide identifier regions to obtain codons attached to the bifunctional complexes having received the chemical entities,

f) separating the components of the hybridisation products and recovering the complexes,

g) repeating steps c) to f) one or more times, as appropriate.

31. The method according to claim 30, wherein the anticodon is unique.

32. The method of claims 30 and 31, wherein a similar combination of nucleotides of an anti-codon does not appear on another building block carrying another functional entity.

33. A method for generating a library of bifunctional complexes comprising a display molecule part and a coding part, said method comprising the steps of providing in separate compartments nascent bifunctional complexes, each comprising a chemical reaction site and a priming site for enzymatic addition of a tag and performing in any order reaction in each compartment between the chemical reaction site and one or more reactants, and addition of one or more respective tags identifying the one or more reactants at the priming site using one or more enzymes.

34. The method according to claim 33, wherein the content of two or more compartments are pooled together.

35. The method of claims 33 or 34, wherein the bifunctional complexes are subjected to one or more further cycles of the method according to claims 33 or 34, using the bifunctional complexes as the nascent bifunctional complexes, to obtain a library of bifunctional complexes, in which each member of the library comprises a reagent specific reaction product and respective tags which codes for the identity of each of the reactants that have participated in the formation of the reaction product.

36. The method according to any of the preceding claims for generating a library of bifunctional complexes, each of the members comprising a display molecule part formed by reaction involving a chemical reaction site and two or more reactants, and a coding part comprising respective codons coding for the identity of the two or more

reactants, comprising the steps of selecting a method according to claims 30-32 or 33-35 and using the obtained premature library of bifunctional complexes as a pool of different nascent bifunctional complexes in a method selected from the method according to claim 30-32 and the method according to claim 33-35.

37. A method for identifying a display molecule having a preselected property, comprising the steps of:

i) subjecting the library formed in accordance with any of the claims 30 to 36 to a condition, wherein a display molecule or a subset of display molecules having a predetermined property is partitioned from the remainder of the library, and

ii) identifying the display molecule(s) having a preselected property by decoding the encoding part of the complex.

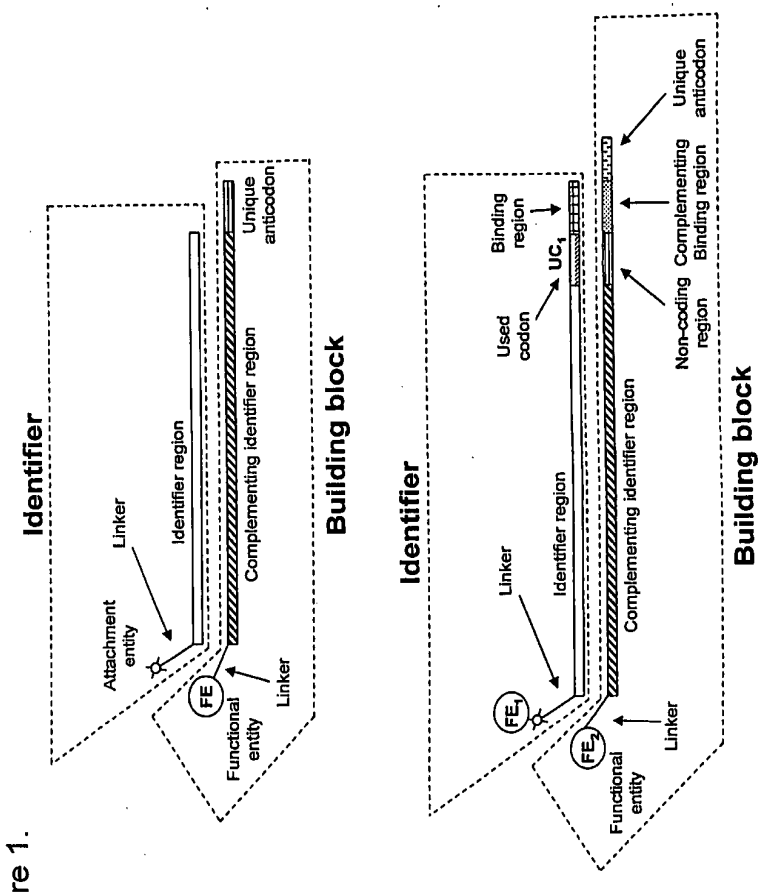


Figure 1.

Figure 2.

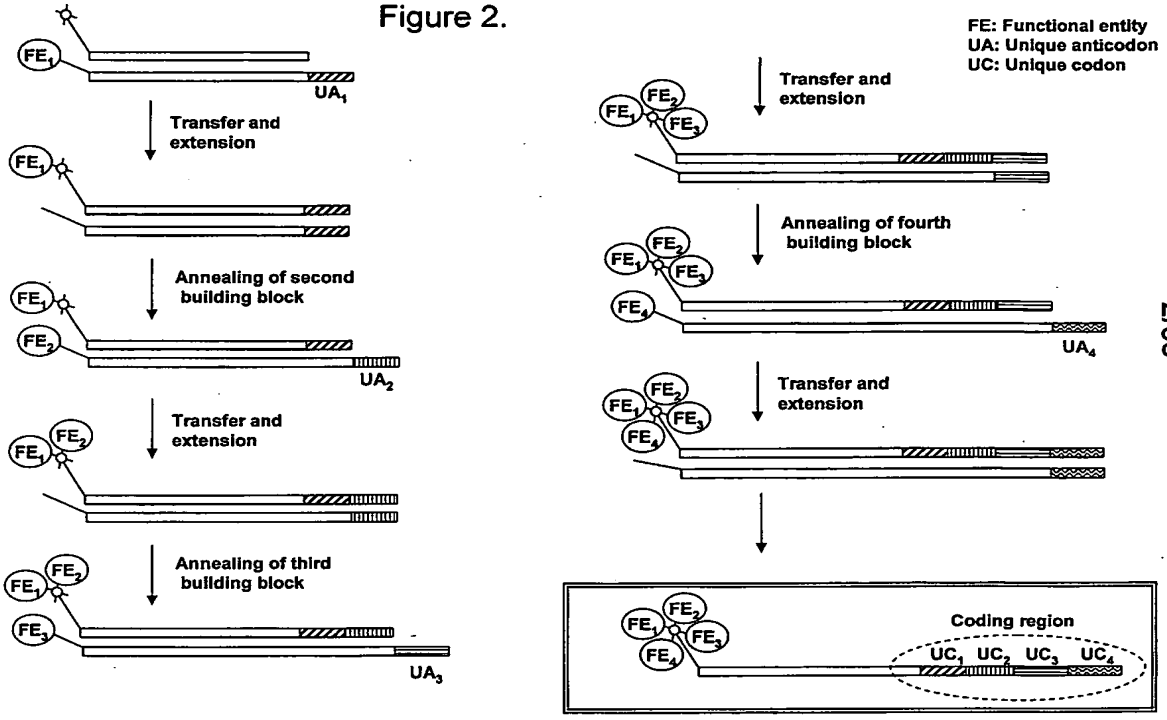
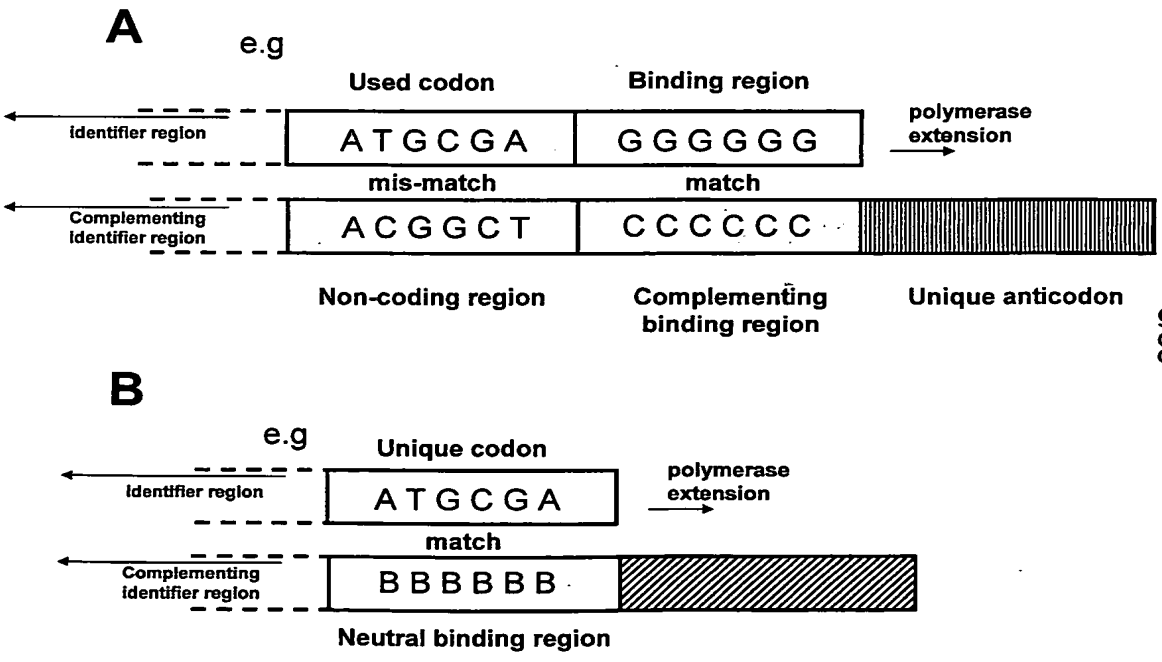


Figure 3.



WO 2004/039825 PCT/DK2003/000739

Figure 4.

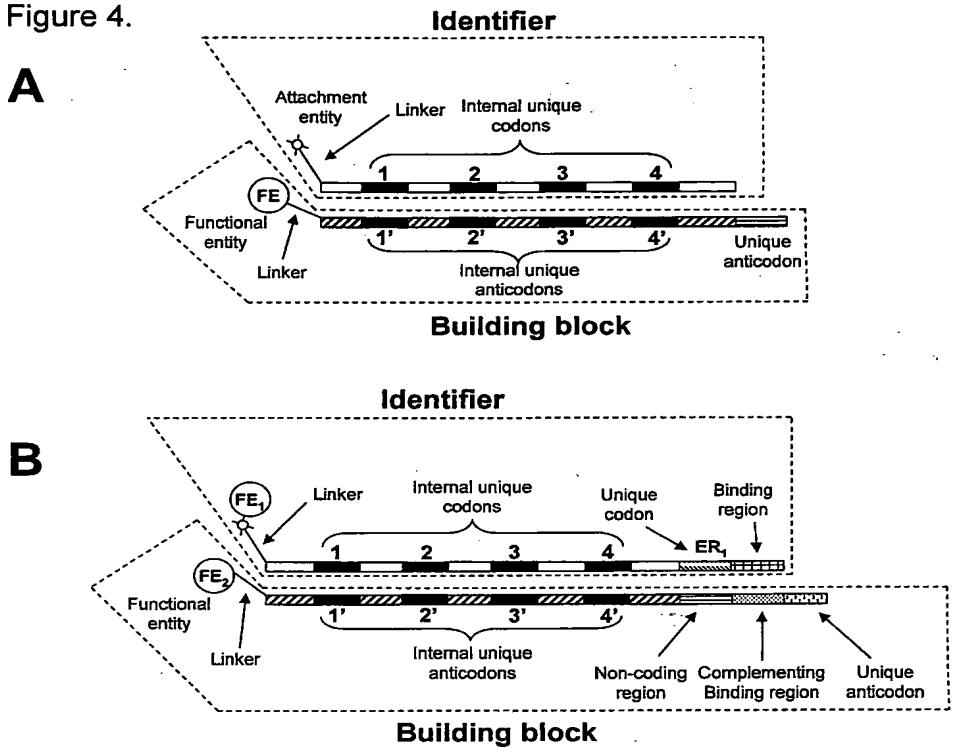


Figure 5.

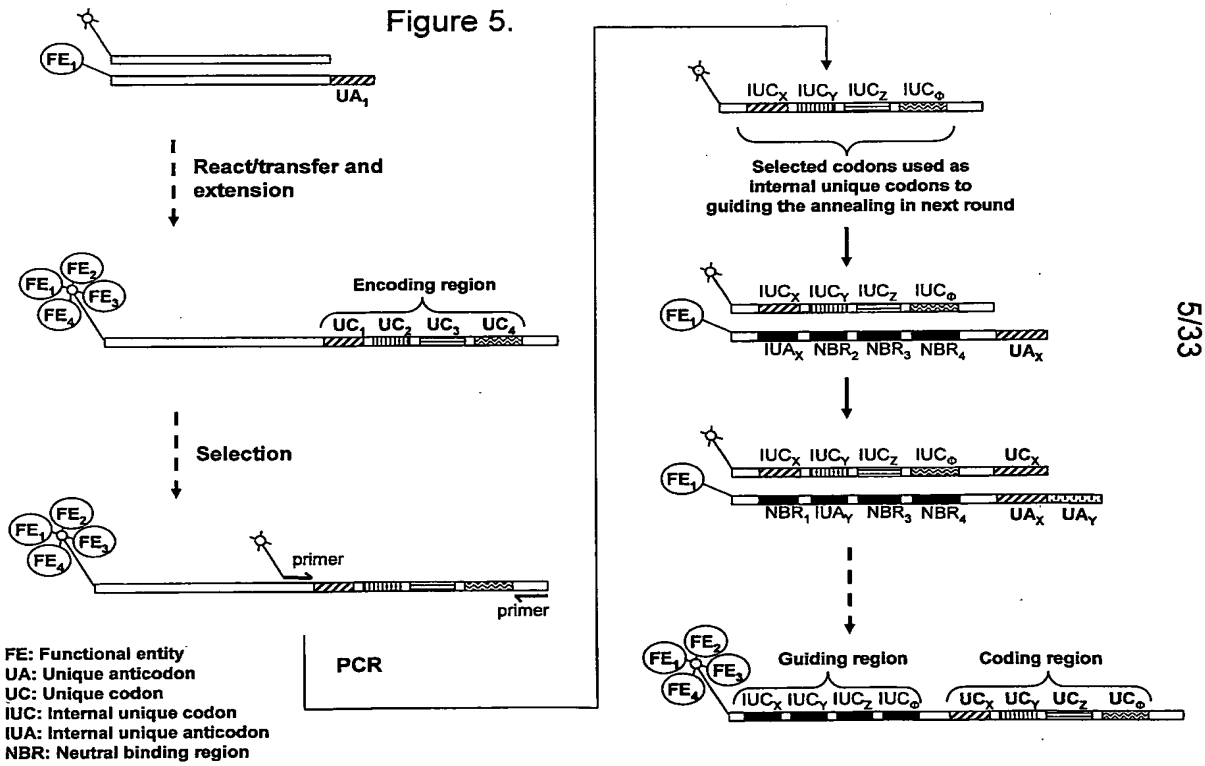


Figure 6.

FE: Functional entity
UA: Unique anticodon
UC: Unique codon
IUC: Internal unique codon

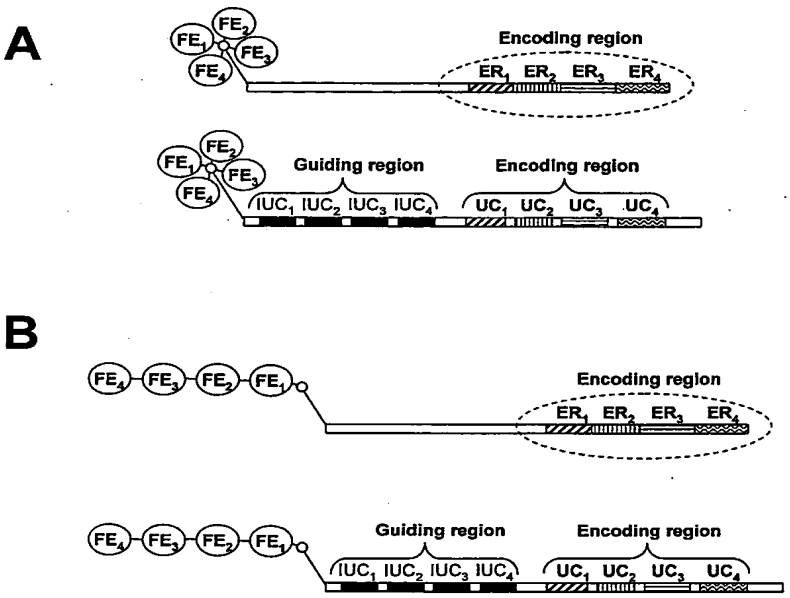
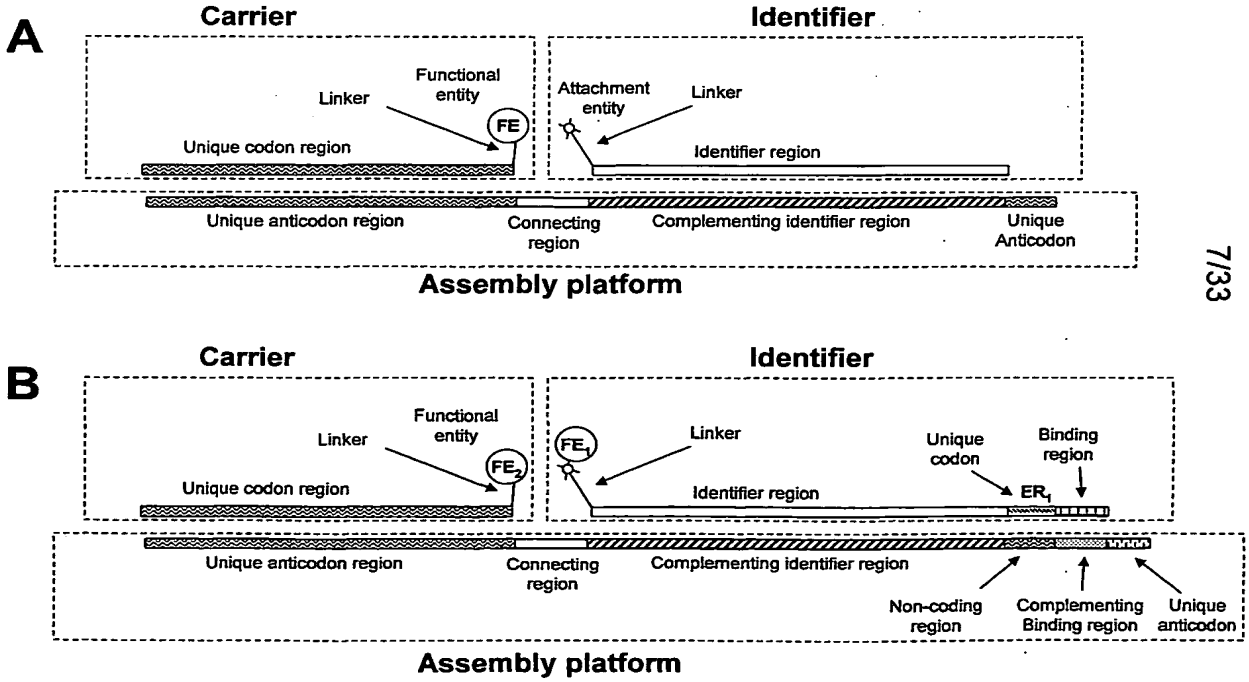


Figure 7.



WO 2004/039825 PCT/DK2003/000739

Figure 8.

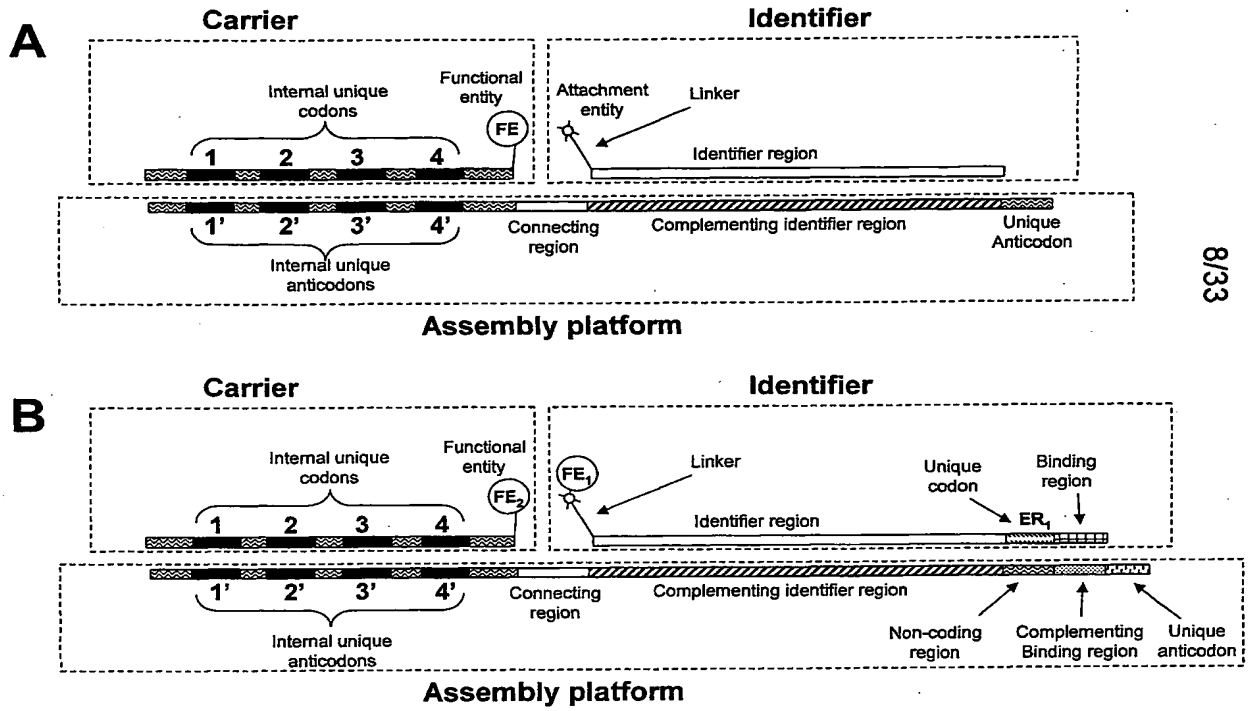


Figure 9.

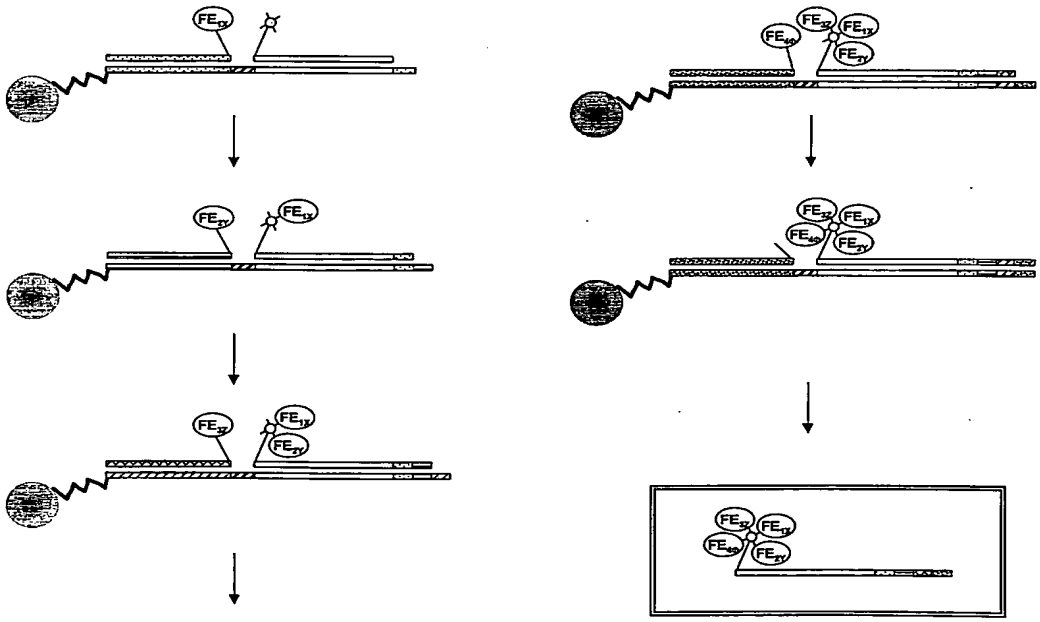
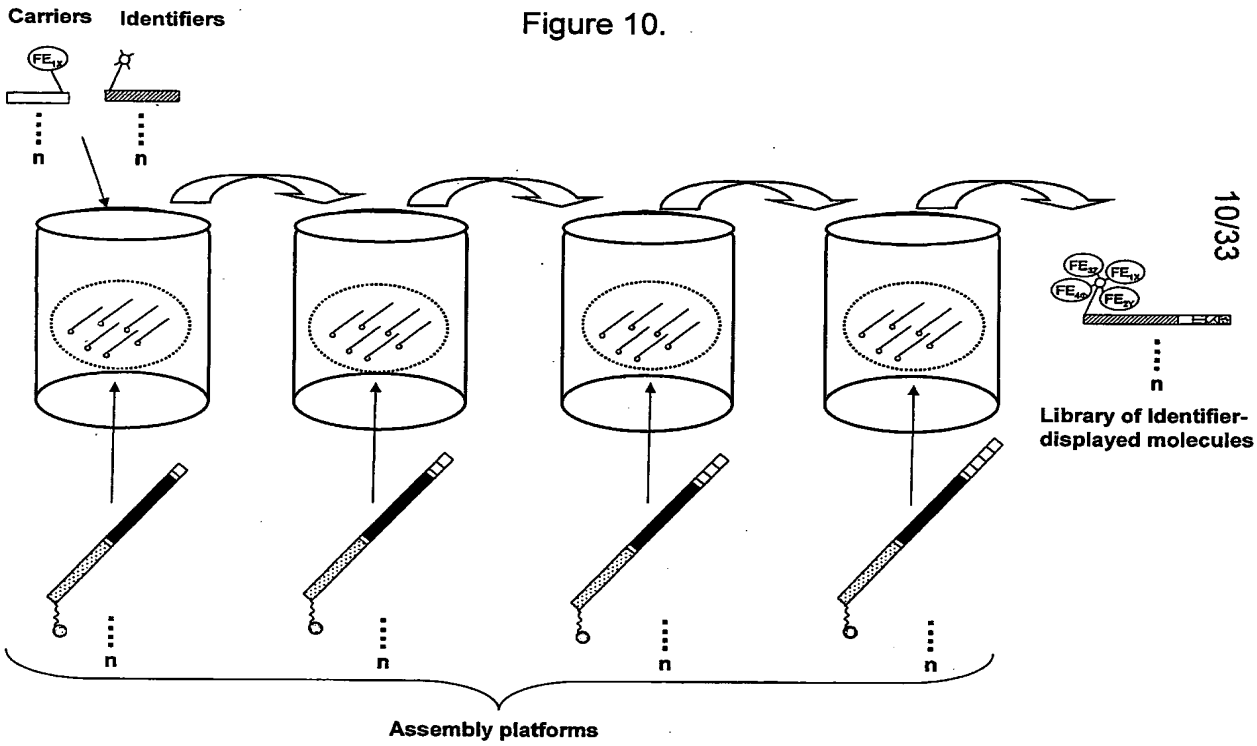
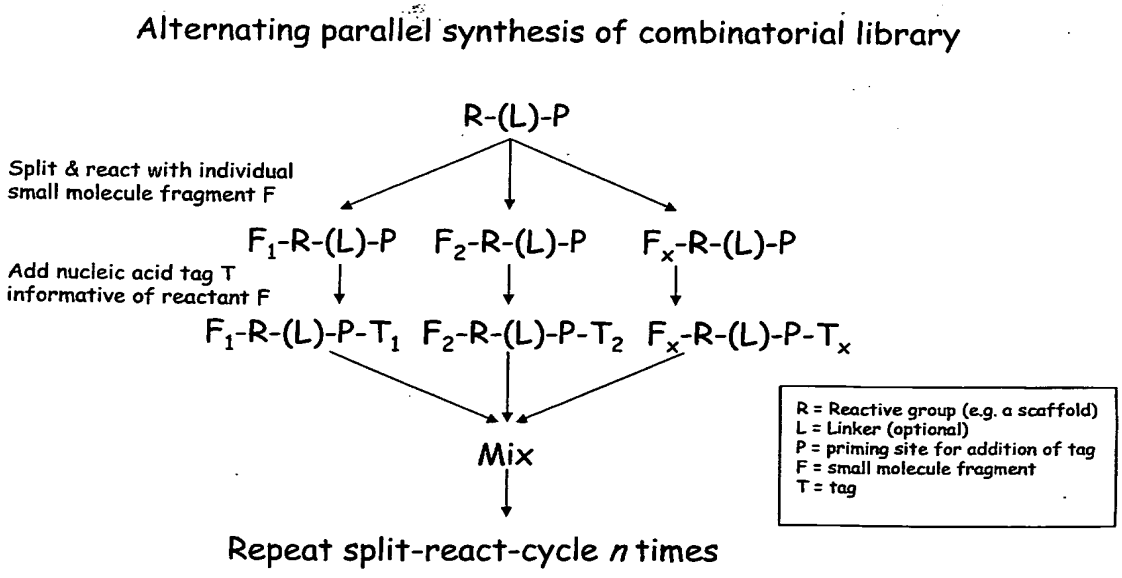


Figure 10.



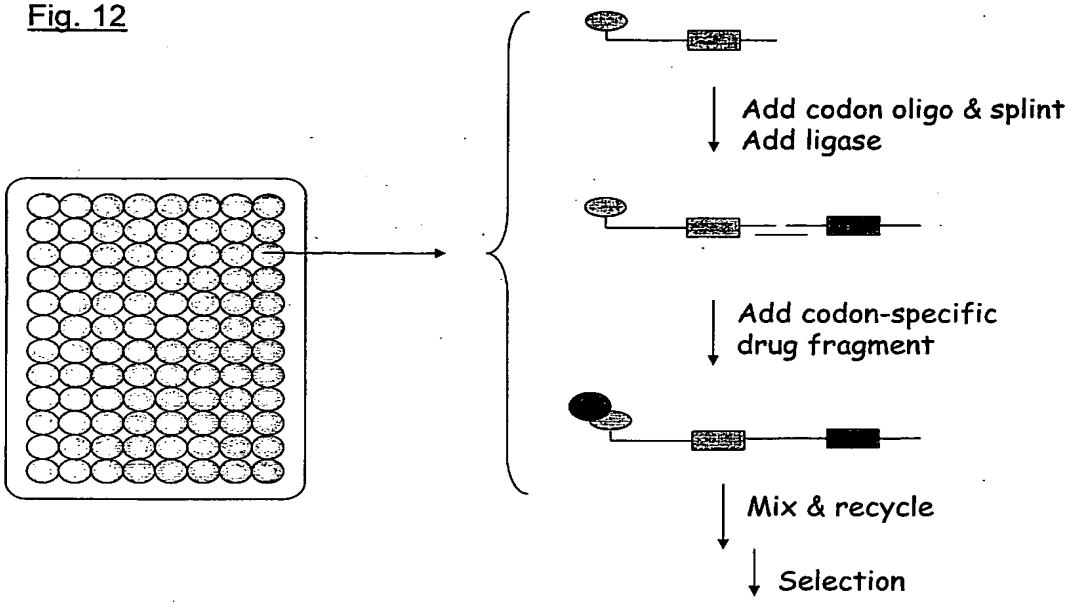
WO 2004/039825 PCT/DK2003/000739

Fig. 11



WO 2004/039825 PCT/DK2003/000739

Fig. 12



WO 2004/039825 PCT/DK2003/000739

Fig. 13

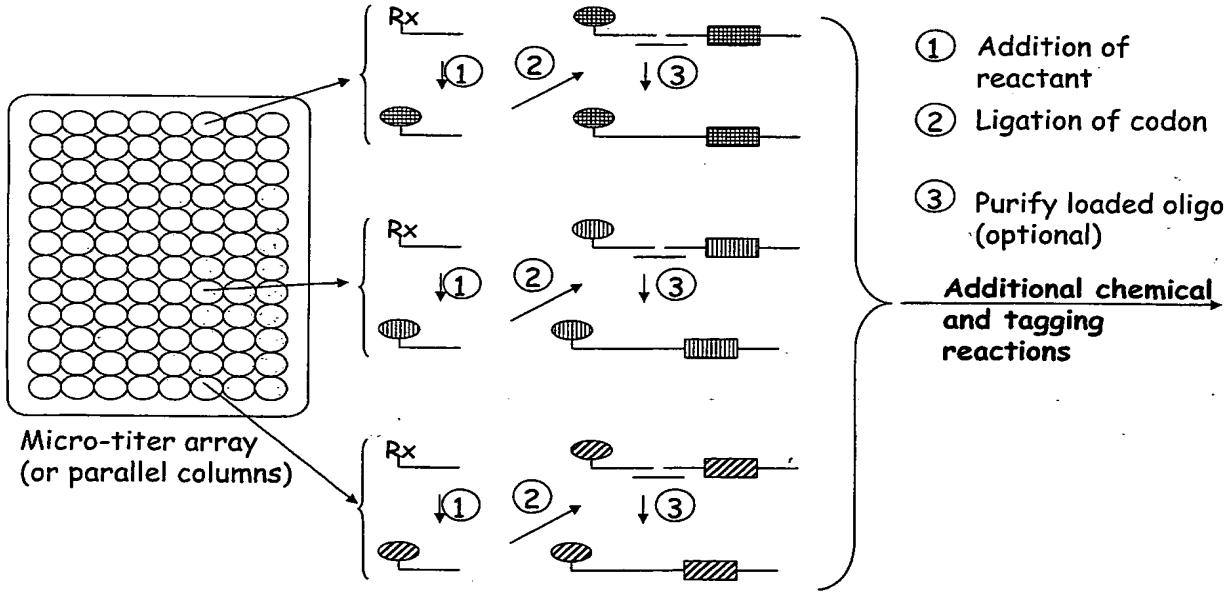


Fig. 14

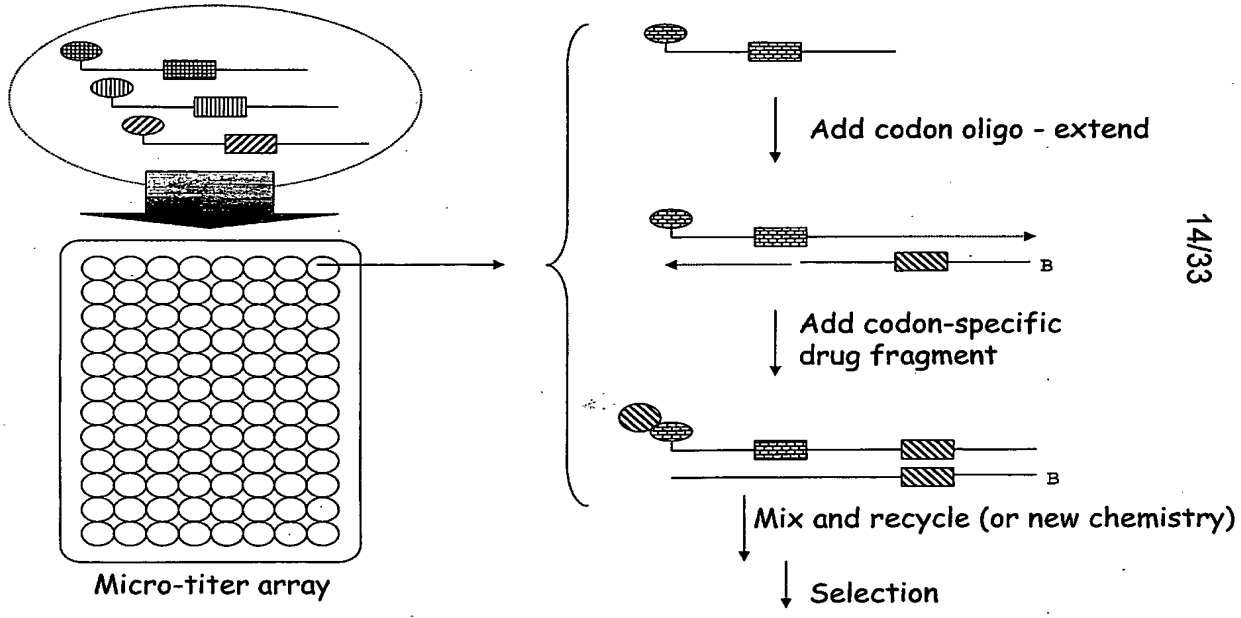
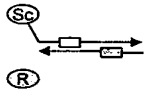
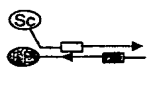
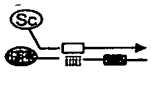
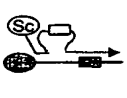
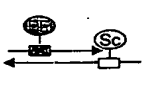
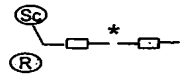
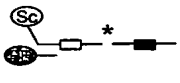
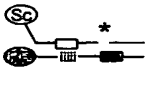
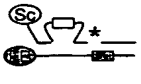
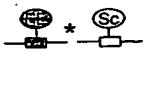
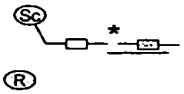
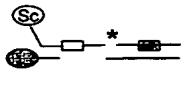
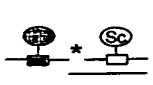
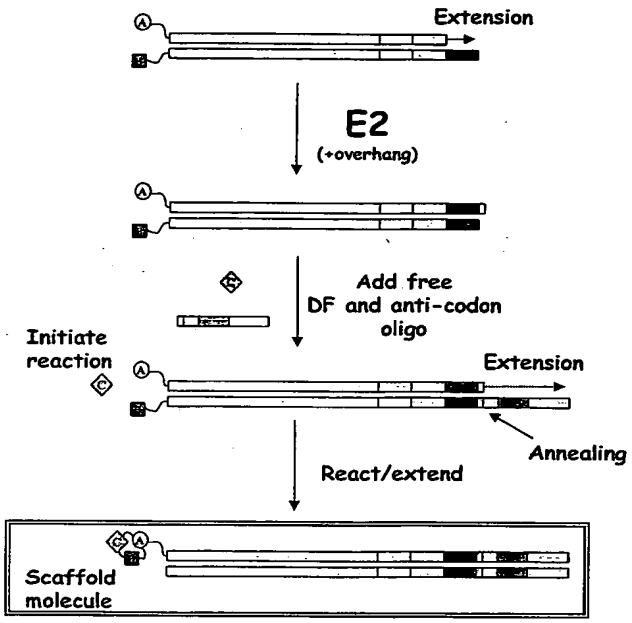


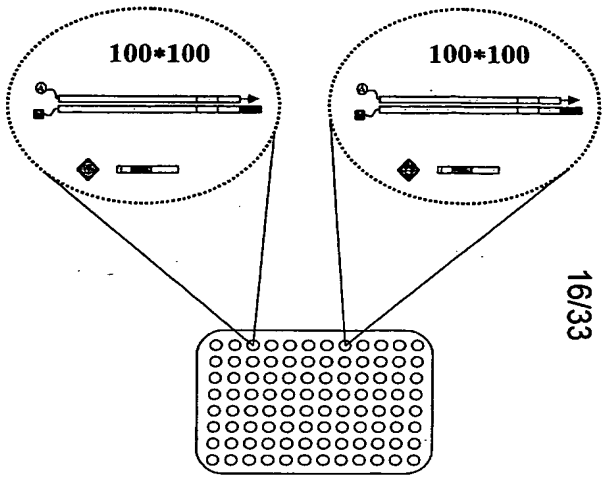
Fig. 15 Single encoding

Reactant / Encoding method	Free Reactant	Zipper Building Block	E2 Building block	Loop Building block	N Building Block
Polymerase					
Ligase (ss)					
Ligase (ds)					

WO 2004/039825 PCT/DK2003/000739

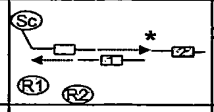
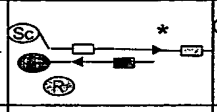
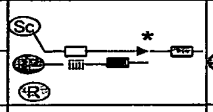
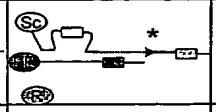
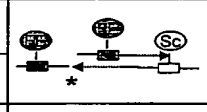
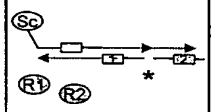
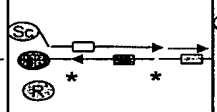
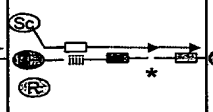
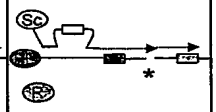
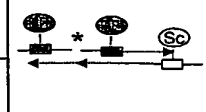
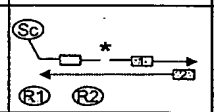
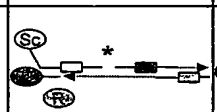
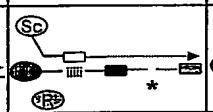
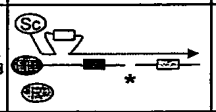
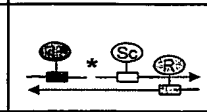
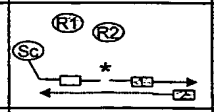
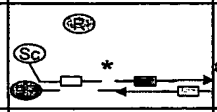
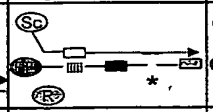
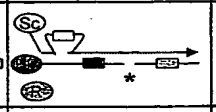
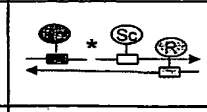
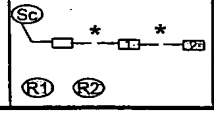
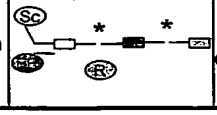
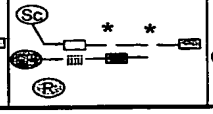
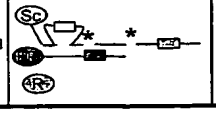
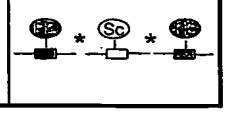
Fig. 16





10^4 different molecules generated in each well. 100 wells gives 10^6 different molecules

Fig. 17 Double encoding

Reactant / Encoding method	Free Reactant	Zipper Building Block	E2 Building block	Loop Building block	N Building Block
Polymerase extension then single stranded ligation[1]					
Polymerase extension then double stranded ligation[1]					
Single stranded ligation[1] then polymerase extension					
Double stranded ligation[1] then polymerase extension					
Ligation then ligation[2]					

[1] Enzymatic or chemical ligation

[2] Single or double stranded enzymatic or chemical ligation

WO 2004/039825 PCT/DK2003/000739

Fig. 18

Fig. 18

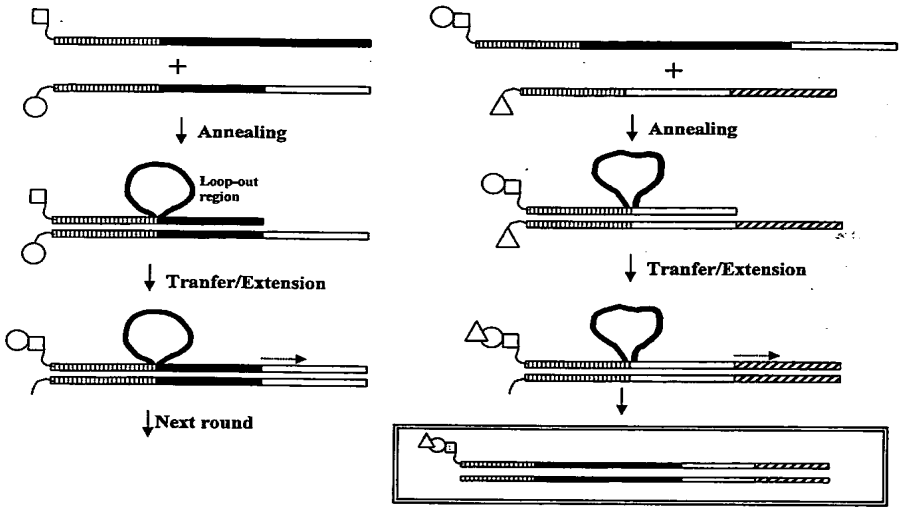


Fig. 19

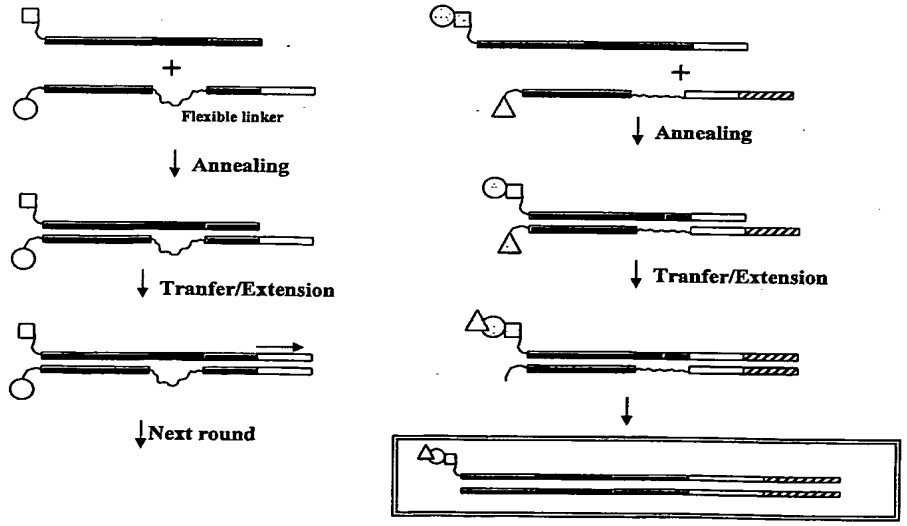


Fig. 20

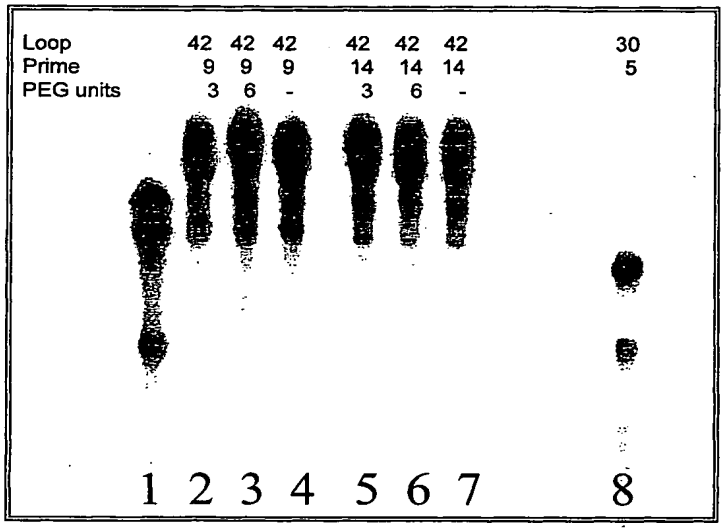
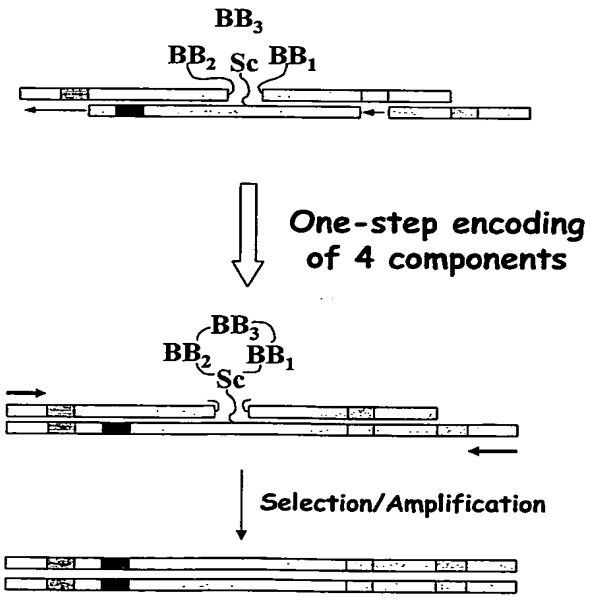

Loop
Prime
PEG units
42 42 42 42 42 42 30
9 9 9 14 14 14 5
3 6 - 3 6 -
1 2 3 4 5 6 7 8

Fig. 21


BB3
BB2 Sc BB1
One-step encoding of 4 components
BB3
BB2 BB1
Sc
Selection/Amplification

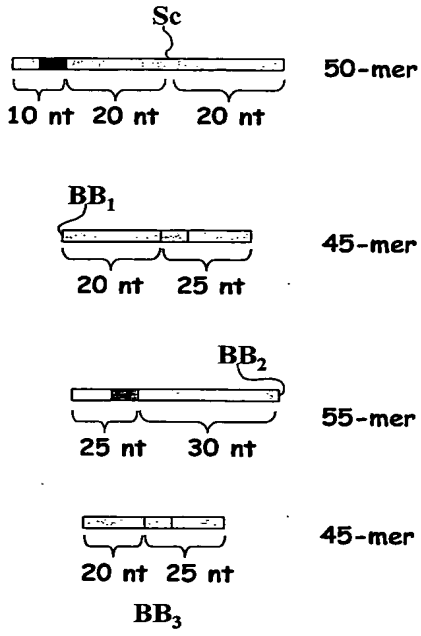

Sc
10 nt 20 nt 20 nt
50-mer
BB1
20 nt 25 nt
45-mer
BB2
25 nt 30 nt
55-mer
20 nt 25 nt
BB3
45-mer

Fig. 22

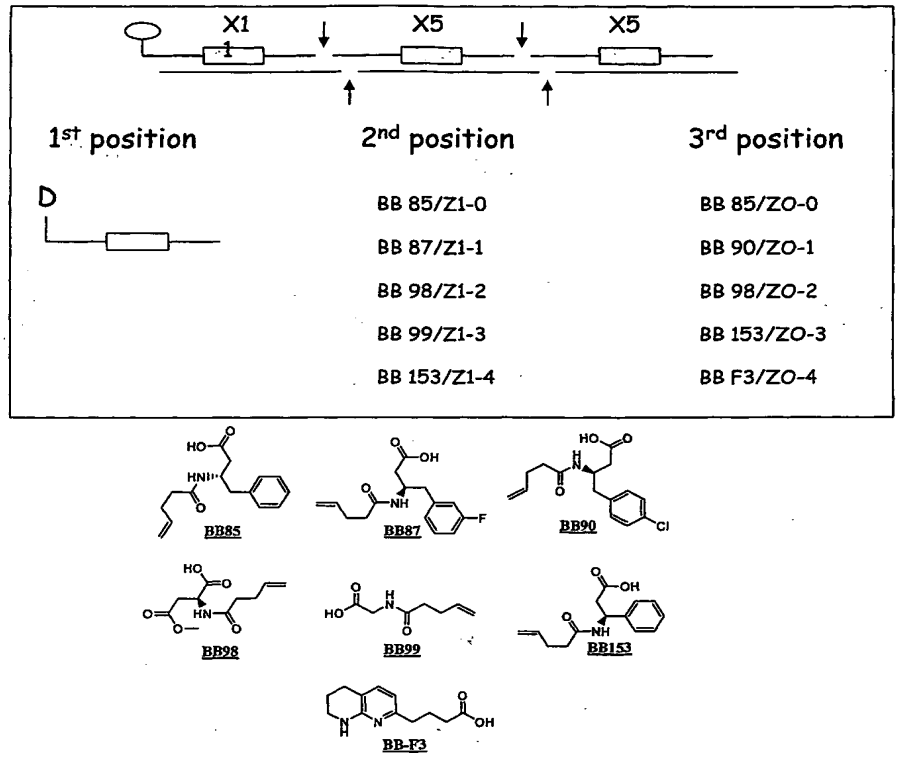

X1
1
X5
X5
1st position
2nd position
3rd position
D
BB 85/Z1-0
BB 87/Z1-1
BB 98/Z1-2
BB 99/Z1-3
BB 153/Z1-4
BB 85/ZO-0
BB 90/ZO-1
BB 98/ZO-2
BB 153/ZO-3
BB F3/ZO-4
BB85
BB87
BB90
BB98
BB99
BB153
BB-F3

Fig. 23

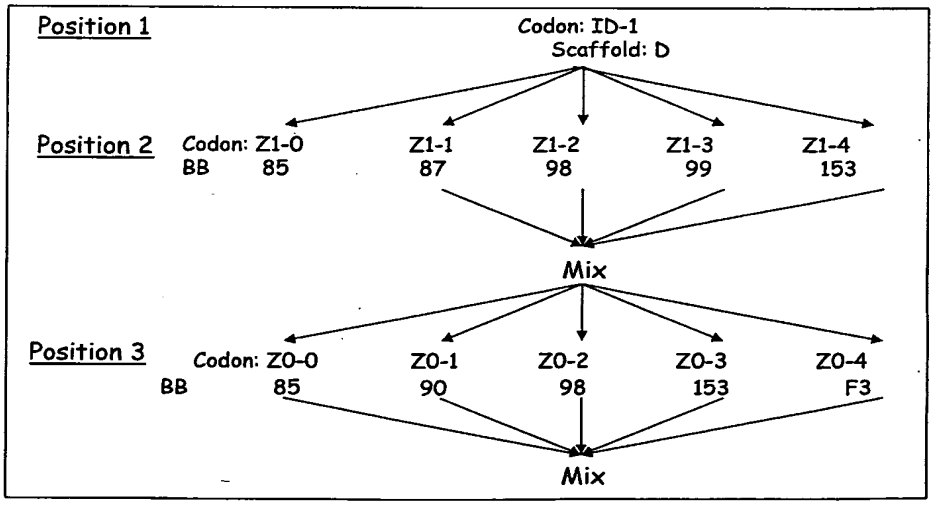

Position 1
Codon: ID-1
Scaffold: D
Position 2
Codon: Z1-0
BB 85
Z1-1
87
Z1-2
98
Z1-3
99
Z1-4
153
Mix
Position 3
Codon: ZO-0
BB 85
ZO-1
90
ZO-2
98
ZO-3
153
ZO-4
F3
Mix

Fig. 24

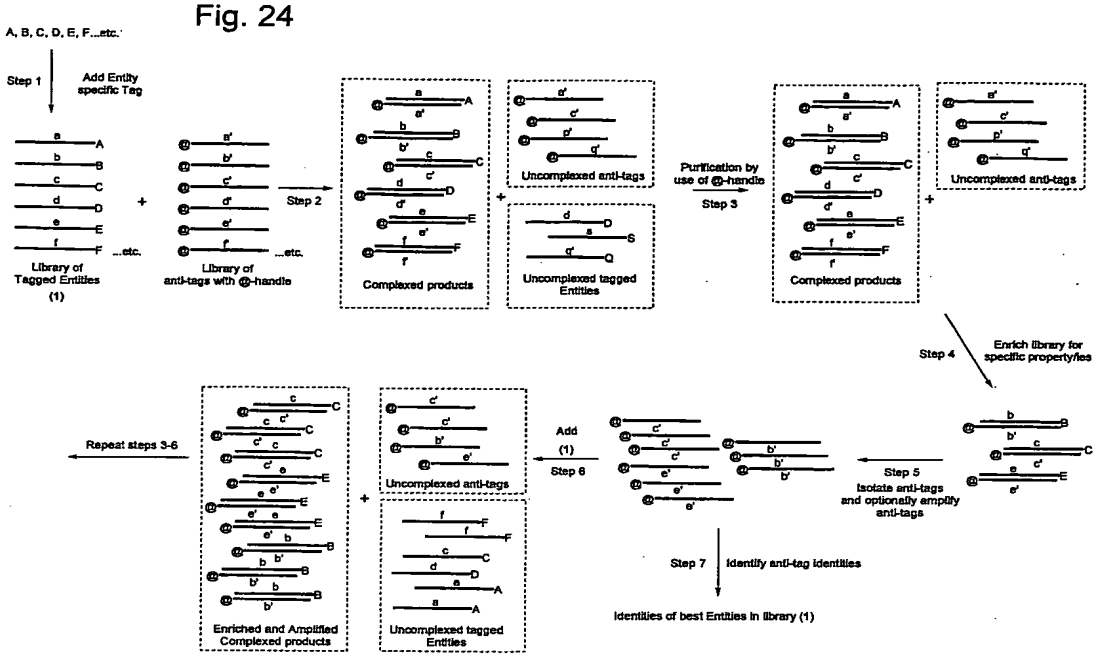
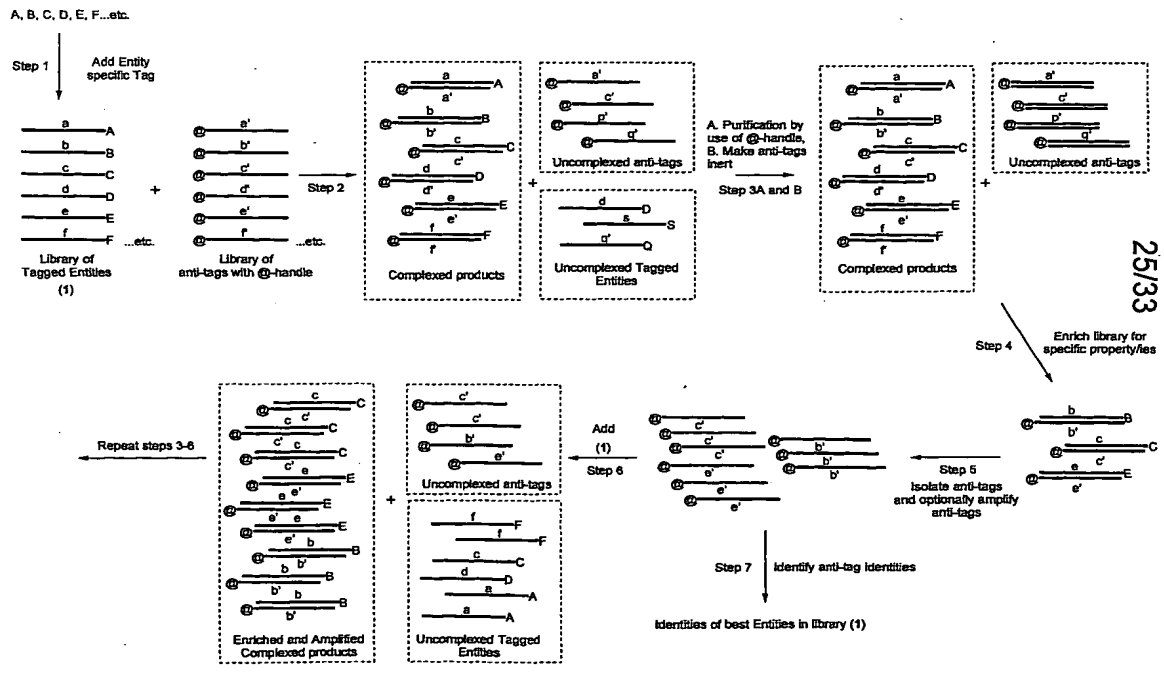


Fig. 25



WO 2004/039825 PCT/DK2003/000739

Fig. 26

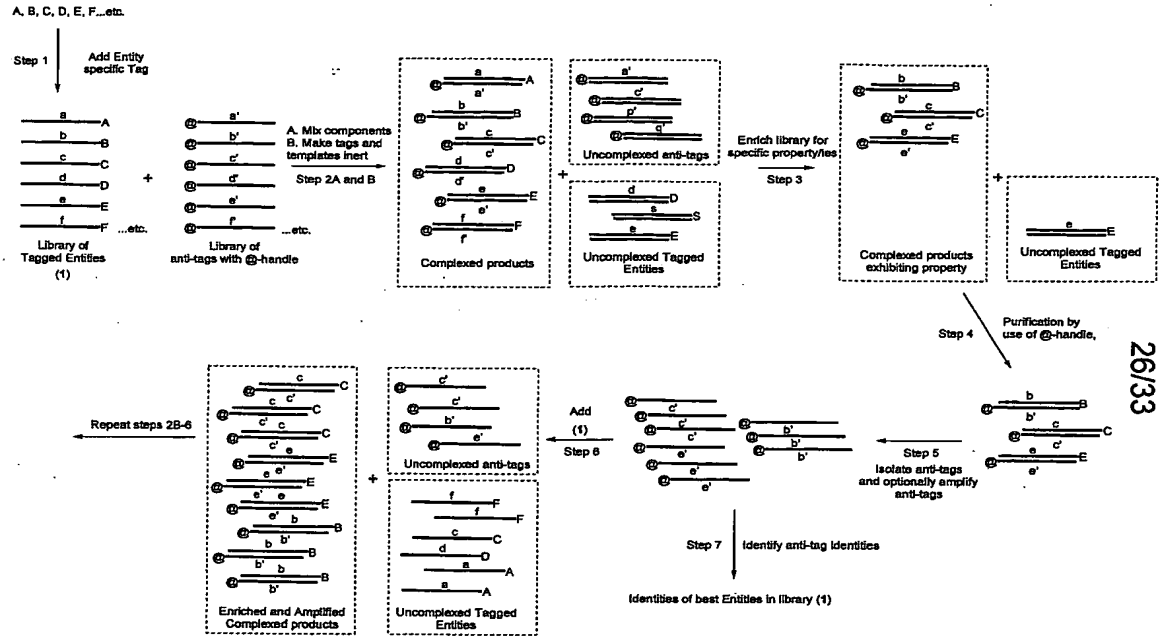
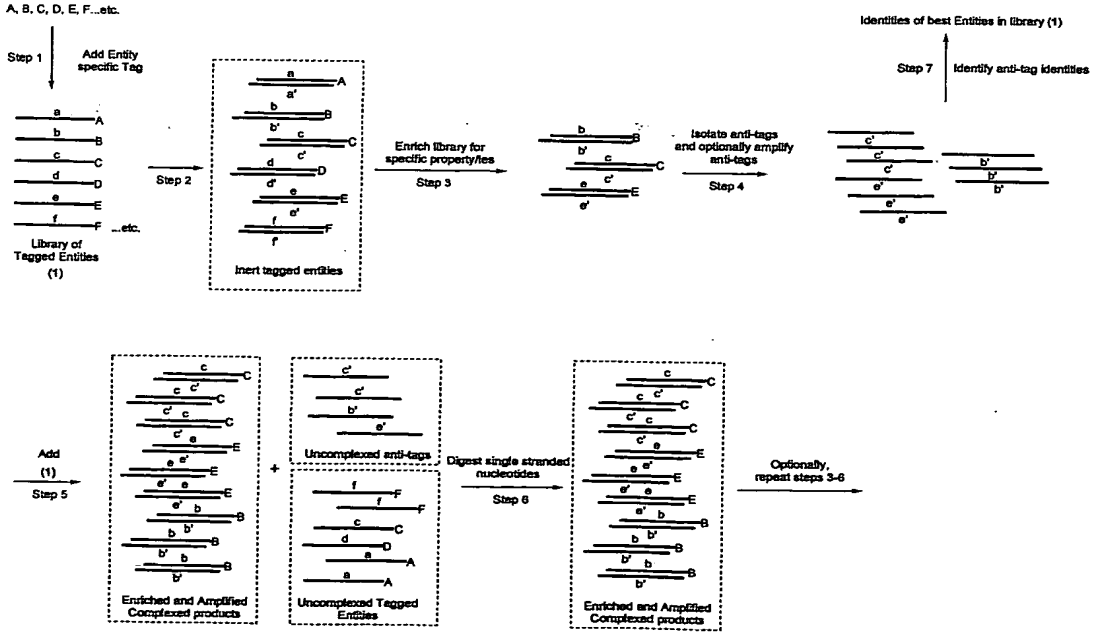


Fig. 27



WO 2004/039825 PCT/DK2003/000739

Fig. 28

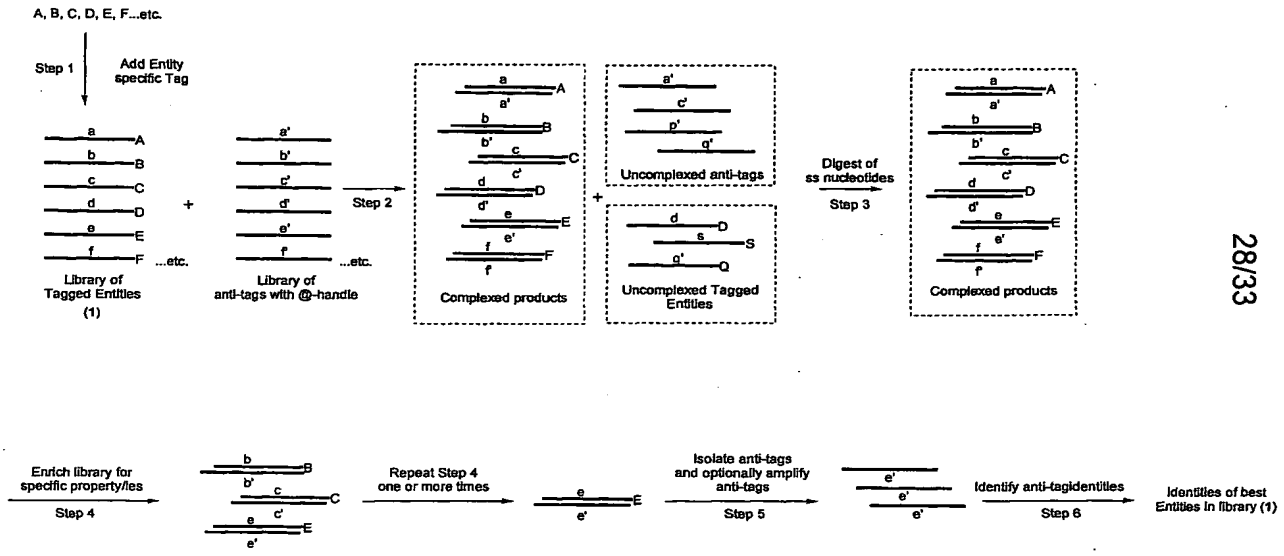

A, B, C, D, E, F...etc.
Step 1
Add Entity specific Tag
Library of Tagged Entities (1)
Library of anti-tags with @-handle
Step 2
Complexed products
Uncomplexed anti-tags
Uncomplexed Tagged Entities
Digest of ss nucleotides
Step 3
Complexed products
Enrich library for specific property/ies
Step 4
Repeat Step 4 one or more times
Isolate anti-tags and optionally amplify anti-tags
Step 5
Identify anti-tagidentities
Step 6
Identities of best Entities in library (1)

Fig. 29

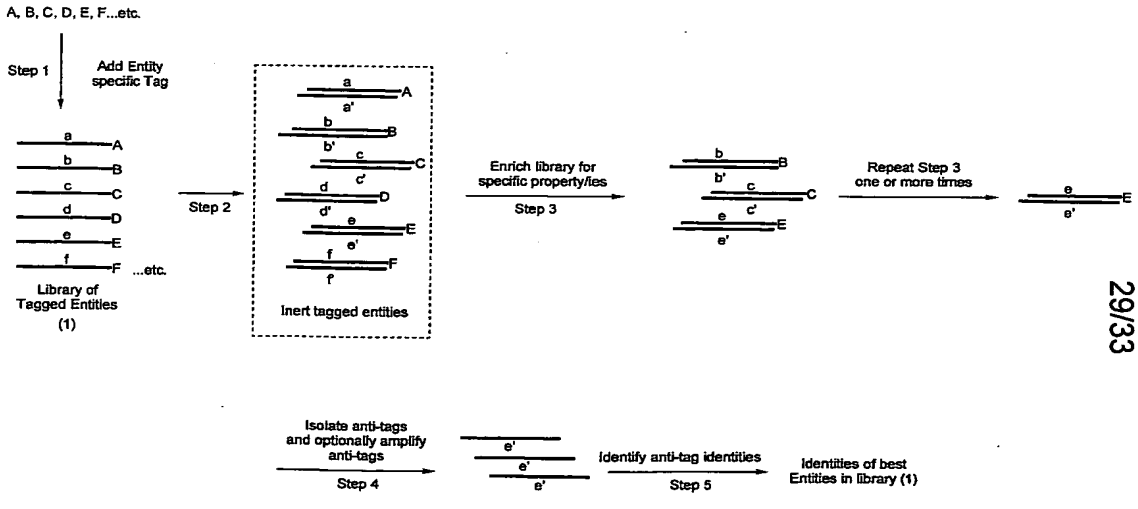

A, B, C, D, E, F...etc.
Step 1
Add Entity specific Tag
Library of Tagged Entities (1)
Step 2
Inert tagged entities
Enrich library for specific property/ies
Step 3
Repeat Step 3 one or more times
Isolate anti-tags and optionally amplify anti-tags
Step 4
Identify anti-tag identities
Step 5
Identities of best Entities in library (1)

Fig. 30

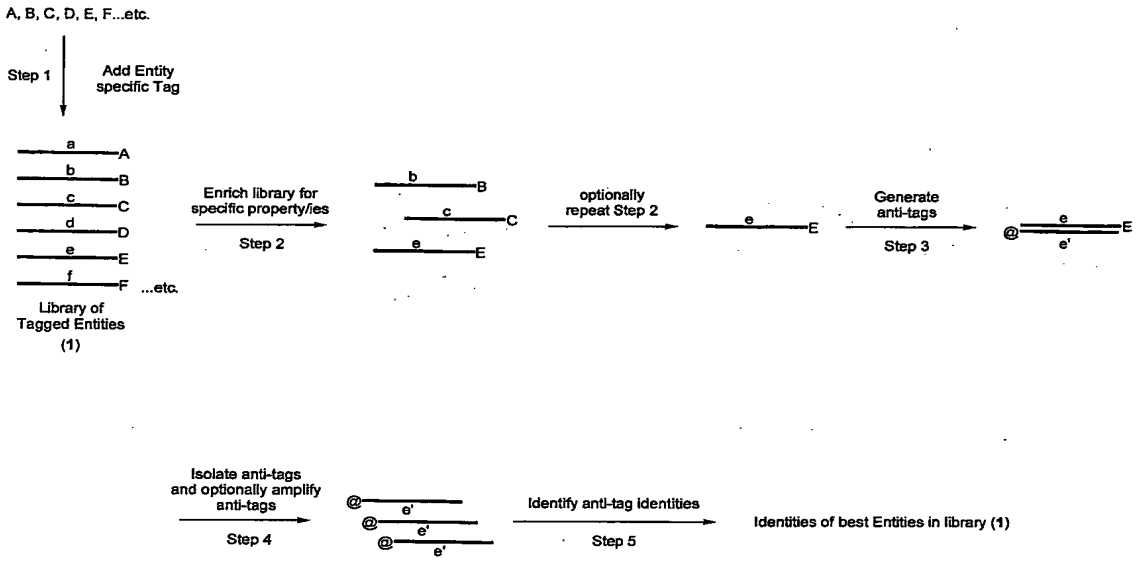
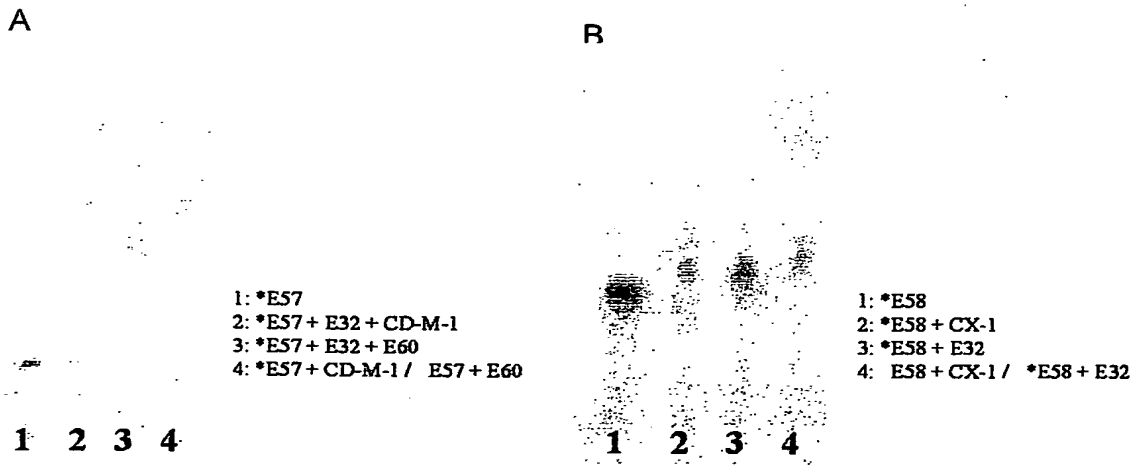


Fig. 31



WO 2004/039825 PCT/DK2003/000739

Fig. 32

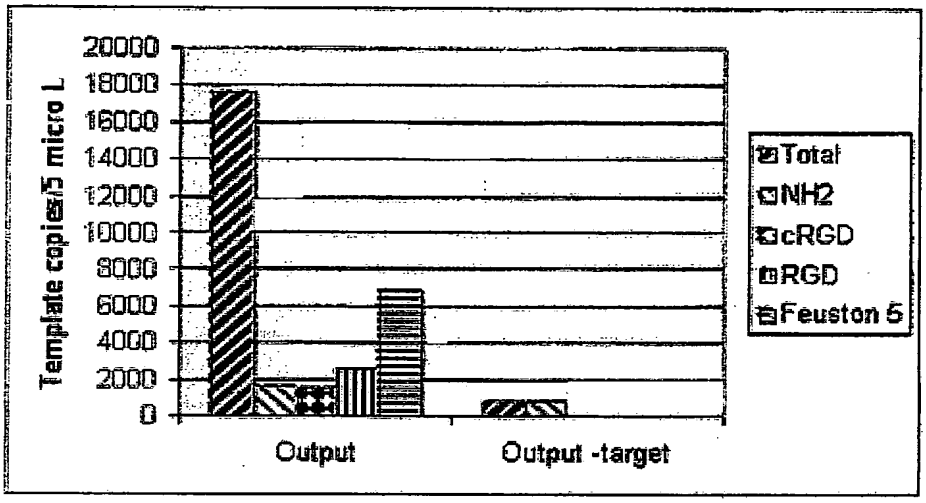
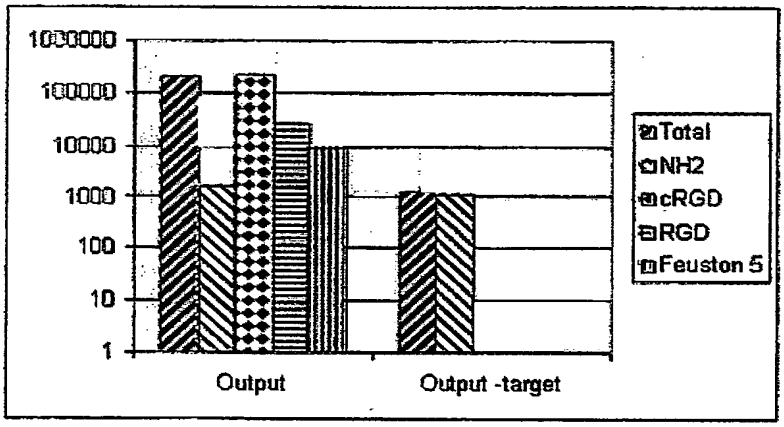

Template copies/5 micro L
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Output
Output -target
Total
NH2
cRGD
RGD
Feuston 5

Fig. 33


1000000
100000
10000
1000
100
10
1
Output
Output -target
Total
NH2
cRGD
RGD
Feuston 5

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 May 2004 (13.05.2004)

PCT

(10) International Publication Number
WO 2004/039825 A3

(51) International Patent Classification[7]: C12P 19/34, C07B 61/00

(21) International Application Number: PCT/DK2003/000739

(22) International Filing Date: 30 October 2003 (30.10.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

PA 2002 01652	30 October 2002 (30.10.2002)	DK
60/422,167	30 October 2002 (30.10.2002)	US
PA 2002 01955	19 December 2002 (19.12.2002)	DK
60/434,425	19 December 2002 (19.12.2002)	US
PA 2003 01064	11 July 2003 (11.07.2003)	DK
60/486,199	11 July 2003 (11.07.2003)	US

(71) Applicant *(for all designated States except US)*: **NUEVOLUTION A/S** [DK/DK]; Rønnegade 8, 5th floor, DK-2100 Copenhagen Ø (DK).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FRESKGÅRD, Per-Ola** [SE/SE]; Örtugsgatan 50, S-603 79 Norrköping (SE). **FRANCH, Thomas** [DK/DK]; Humlebækgade 14, st, tv., DK-2200 København N (DK). **GOULIAEV, Alex, Haahr** [DK/DK]; Brøndsted 223, DK-3670 Veksø Sjælland (DK). **LUNDORF, Mikkel, Dybro** [DK/DK]; Viborggade 13, st. tv., DK-2100 Cpèenjagen Ø (DK). **FELDING, Jakob** [DK/DK]; Ordruphøjvej 24, 1., DK-2920 Charlottenlund (DK). **OLSEN, Eva, Kampmann** [DK/DK]; Vingetoften 17, DK-2730 Helev (DK). **HOLTMANN, Anette** [DK/DK]; Langekærvej 42, DK-2750 Ballerup (DK). **JAKOBSEN, Søren, Nyboe** [DK/DK]; Frederiksvej 36, 4. tv., DK-2000_Frederiksberg (DK). **SAMS, Christian** [DK/DK]; Jakob Dannefaerds Vej 4a 1, DK-1973 Frederiksberg C (DK). **GLAD, Sanne, Schrøder** [DK/DK]; Viggo Barfoeds Allé 59, DK-2750 Ballerup (DK). **JENSEN, Kim, Birkebæk** [DK/DK]; Voldumvej 30C, DK-2610 Rødovre (DK). **PEDERSEN, Henrik** [DK/DK]; Frodesvej 24, DK-2880 Bagsvaerd (DK).

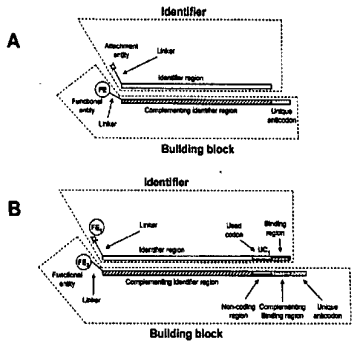
(74) Agent: **HOIBERG A/S**; Store Kongensgade 59 A, DK-1264 Copenhagen K. (DK).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD FOR THE SYNTHESIS OF A BIFUNCTIONAL COMPLEX



(57) Abstract: A method for synthesising a bifunctional complex which comprises a reaction product and an encoding region. The complex is a hybridisation product between a nascent bifunctional complex and a building block. The nascent bifunctional complex comprises an attachment entity linked to an oligonucleotide identifier region and the building block comprises a fuctional entity attached to an oligonucleotide which is sufficiently complementary to the identifier region to allow for hybridisation . The functional entity may then be transferred to the attachment entity by a chemical reaction.The complementary identifier region further comprises a unique codon, which identifies the functional entity unequivocally. This codon functions as a template for forming a unique codon as an extention of the identifier region by use of an enzime, such as a polymerase. The unique codon may be used to decode the synthetic history of the reaction product after several cycles of transfer of functional entities.

WO 2004/039825 A3

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
28 October 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

INTERNATIONAL SEARCH REPORT

Int al Application No
PCT/DK 03/00739

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12P19/34 C07B61/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12P C07B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 02 074929 A (KANAN MATTEW W;GARTNER ZEV J ; LIU DAVID R (US); HARVARD COLLEGE ()) 26 September 2002 (2002-09-26) cited in the application paragraph [0006] - paragraph [0010] paragraph [0106]; claims; figures 1-57 ---	1-37
E	WO 2004 024929 A (FRANCH THOMAS ;NUEVOLUTION AS (DK); THISTED THOMAS (DK)) 25 March 2004 (2004-03-25) figures 1-25 ---	1-37
E	WO 2004 013070 A (NUEVOLUTION AS ;PEDERSEN HENRIK (DK)) 12 February 2004 (2004-02-12) page 52, paragraph 2; figures 1-60 ---	1-37
	-/--	

[X] Further documents are listed in the continuation of box C. [X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 June 2004	30 AUG 2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	IDA CHRISTENSEN / ELY

INTERNATIONAL SEARCH REPORT

Int Application No
PCT/DK 03/00739

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SUMMERER D ET AL: "DNA-templated synthesis: more versatile than expected" ANGEWANDTE CHEMIE, WILEY-VCH, WEINHEIM, DE, vol. 41, no. 1, 4 January 2002 (2002-01-04), pages 89-90, XP002265218 ISSN: 1433-7851 the whole document ---	1-37
A	WO 94 08051 A (READER JOHN C ;STILL W CLARK (US); UNIV COLUMBIA (US); COLD SPRING) 14 April 1994 (1994-04-14) abstract -----	1-37

INTERNATIONAL SEARCH REPORT

information on patent family members

Inte al Application No
PCT/DK 03/00739

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 02074929	A	26-09-2002	US	2003113738 A1	19-06-2003
			CA	2441820 A1	26-09-2002
			EP	1423400 A2	02-06-2004
			WO	02074929 A2	26-09-2002
WO 2004024929	A	25-03-2004	WO	03078625 A2	25-09-2003
			WO	2004024929 A2	25-03-2004
WO 2004013070	A	12-02-2004	WO	2004013070 A2	12-02-2004
WO 9408051	A	14-04-1994	AT	244769 T	15-07-2003
			AU	716621 B2	02-03-2000
			AU	4525897 A	12-02-1998
			AU	686579 B2	12-02-1998
			AU	5536994 A	26-04-1994
			CA	2143848 A1	14-04-1994
			DE	69333087 D1	14-08-2003
			DE	69333087 T2	06-05-2004
			DK	665897 T3	20-10-2003
			EP	0665897 A1	09-08-1995
			ES	2204910 T3	01-05-2004
			HU	72495 A2	28-05-1996
			IL	107166 A	31-10-2000
			JP	8506175 T	02-07-1996
			NO	951230 A	30-03-1995
			PT	665897 T	28-11-2003
			WO	9408051 A1	14-04-1994
			US	5565324 A	15-10-1996
			US	5968736 A	19-10-1999
			US	5721099 A	24-02-1998
			US	5789172 A	04-08-1998
			US	6503759 B1	07-01-2003